DIVISION OF AGRICULTURAL SCIENCES
UNIVERSITY OF CALIFORNIA

ANALYSES OF CALIFORNIA FARM INCOME RELATIONSHIPS

Gary Elsner and Irving Hoch

CALIFORNIA AGRICULTURAL EXPERIMENT STATION
GIANNINI FOUNDATION OF AGRICULTURAL ECONOMICS

Giannini Foundation Research Report No. 297
August 1968

ACKNOWLEDGMENTS

This report is a revised and extended statement of the core of Gary Elsner's Ph.D. thesis. Elsner developed the major themes and thrust of the work, organized the data, and assembled and developed computer programs used in the analysis. Irving Hoch served as thesis committee chairman, contributed to the over-all organization of the project and to the solution of technical problems, and assumed major responsibility for the writing of this report. The authors received helpful technical advice and criticism from H. D. Claxton, A. de Janvry, Robert M. Russell, David Sharpnack, and Michael B. Teitz. Professor Raymond G. Bressler, Jr., as was his wont, made especially helpful contributions, both as member of the thesis committee and as reviewer of a draft of this report.

The administrative support of L. L. Sammet and D. A. Clarke, Jr., is acknowledged. Clerical help was given by Nina Bienhoff Johnson, David Foote, Lucille Goodwin, and Emilie Lachmann. Typing and final editing of this report was carried out by Janet Hauge and Ikuko Takeshita.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDIX TABLES

LIST OF FIGURES

ANALYSES OF CALIFORNIA FARM INCOME RELATIONSHIPS

by

Gary Elsner[1/] and Irving Hoch[2/]

Introduction

This study is concerned with the investigation of relationships for rural farm income; this includes examination of the spatial configuration of income and its changes over time, a comparison of farm and nonfarm income over time, and the relation of income distributions to explanatory variables. In the process of this investigation, a number of methodological tools have been forged; it is hoped that these will have further application.

There are two major phases in this study. They can be summarized as follows: (1) investigation of regional patterns of farm income on the basis of county averages and (2) development of a multivariate income distribution determination model, using county income distributions.

In the first phase, average incomes by county were calculated from data appearing in the 1950 and 1960 Census of Population.

A computer program was then developed to organize and plot the county data on a map of California. The program organized the average income data into high, medium, and low categories on the basis of well-defined criteria.

1/ Formerly Assistant Specialist in the Agricultural Experiment Station and on the Giannini Foundation, University of California, Berkeley; presently Agricultural Economist with the Pacific Southwest Forest and Range Experiment Station, Berkeley.

2/ Formerly Associate Professor of Agricultural Economics and Associate Agricultural Economist in the Experiment Station and on the Giannini Foundation, University of California, Berkeley; presently with Resources for the Future, Inc., Washington, D. C.

These organized data were then presented in the form of maps exhibiting each category in turn. (The program can be extended to handle a greater number of categories or other spatial entities.) The first phase of this study also included some investigation of the relation of rural farm to nonfarm income over time and some examination of growth in income over time. No marked difference in income between farm and nonfarm sectors emerged at the county level; but marked regional differences exist and persist through time, though some reduction in these income differentials appears to be occurring. Hence, the problem of income disparity seems more a regional than a sectorial problem in the context of California.

In the second phase of this study, a multivariate analysis was carried out for rural farm income distributions in an investigation of underlying factors affecting income.

Data on county income distributions for the year 1949 appear in the 1950 Census of Population and for the year 1959, in the 1960 Census of Population. For simplicity, these sets of data will be referred to hereafter as the 1950 and 1960 data, respectively. For 1950 there were 14 income classes, and for 1960 there were 11 income classes. For each income class then, there were observations on the fraction of income units falling in that class by county. Treating these county observations as observations on a specific dependent variable, 14 equations were developed for the 1950 data, using 14 different dependent variables and the same set of (1950) explanatory variables. In similar fashion, 11 equations were developed for 1960.

Because there was the likelihood of correlation among the disturbance terms of each set of equations, multivariate analysis was employed.

It may be noted that estimates of coefficients turn out to be the same as those obtained in univariate least-squares regression. However, the concern is with the equation system rather than with individual equations. In

particular, there is focus on the behavior of a variable's coefficients over the set of equations rather than on considering individual parameters an equation at a time. Statistical tests employed reflect this focus.

Results were generally encouraging, with "important" explanatory variables behaving as expected. Thus, increasing the level of education increases the number of persons in the higher income classes and reduces the number in the lower income classes. Increased urbanization shifts income distributions in the same fashion, while increasing the distance to the nearest trading center has the reverse effect. Other variables included in the analysis were measures of real estate value per farm, specialization, irrigation, road quality, tenure, age, and average farm size.

Working with components of the income distribution rather than with the average based on those components is intuitively satisfying. Hopefully, results for the distribution should yield more accurate forecasts and may also afford some basis for policy decisions.

In the following sections of the report, each phase summarized here will be described in detail.

Investigation of Regional Patterns of Farm Income

This section reports the first phase of the present study, which was the investigation of regional patterns of farm income. This section considers (1) the data employed, (2) average income estimates obtained, (3) a technique for spatial analysis mapping, and (4) patterns of income and of income growth and comparisons of farm and nonfarm income.

Income Distribution Data

Selected data on the size distribution of income by California counties for certain sectors for 1949 and 1959 are available in the 1950 and 1960 U. S.

Census of Population, respectively.[1] The data appearing here, by counties, were the basic source for this study.[2] Thus, the regional unit was the county. As noted above, the reference dates for the data throughout this study will be 1950 and 1960 for simplicity of exposition.

The unit of observation is either "family" (f) or "family and unrelated individuals" (fu); this latter category is also referred to as "family and unrelated." The sectors are those defined by the Census by residence. Particular sectors will be indicated by the following abbreviations: rural farm (RF); nonfarm (N), consisting of rural nonfarm and urban combined; and All sectors (A), consisting of rural farm and nonfarm combined.

County income distribution data cover only part of the sectors of interest in both the 1950 and 1960 cases. For 1950, county data were available on family and unrelated units for the All and the nonfarm categories and on family units for the All category only. For 1960, county data were available only on family units; further, data on nonfarm units were not included. Some of the 1950 and 1960 data gaps were filled by direct calculation and some by estimation. The array of data which were initially available plus that developed by direct calculation and estimation are summarized.

Direct calculation was used to obtain (1) 1950 family and unrelated rural farm data and (2) 1960 family nonfarm data, which were calculated by subtraction and addition, respectively, of other categories of information. The 1950 family

---

1/ U. S. Bureau of the Census, U. S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 5, California, 1952, 5-438p.

Idem, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963, 6-990p.

2/ The 1960 Census data differ from the 1950 data in several characteristics. The 1950 Census income data are generally based upon a 20-percent sample while the 1960 data are based upon a 25-percent sample. For both years, Census income data include only dollars of income received and do not record imputed income, for example, rental value of owned housing.

Categories of Available and Developed Data on County Income Distributions
1950 and 1960

Year	Family and unrelated (fu)	Family (f)
1950	A: All = URB + RNF + RF N: nonfarm = URB + RNF RF*: rural farm (direct calculation)	A: All N*: nonfarm (estimated) RF*: rural farm (estimated)
1960	Data not developed for this category	A: All N*: nonfarm (direct calculation) RF: rural farm

* Indicates data were developed by direct calculation or by estimation procedure. Direct calculation means data were obtained by subtraction or addition of other categories. Items not indicated by * are available in the Census.

(Meaning of symbols not defined above: URB = urban; RNF = rural nonfarm.)

nonfarm and the family rural farm distributions were estimated in more involved fashion. It was assumed that the proportionate relation between corresponding elements of the family and family and unrelated individuals for the 1950 All sector also applied to the 1950 nonfarm and farm sectors. Thus, for the farm sector it was assumed that

$$\frac{RF^{*}\ (f)}{RF\ \ (fu)} = \frac{A\ (f)}{A\ (fu)}$$

for each county and income class. Table 1 exhibits the range and number of income class intervals for county income distributions by sector for 1950 and 1960.

For the 1950 data, all income intervals contain income recipients for the rural sector, with the exception of San Francisco. For 1960, the Census

TABLE 1

Range and Number of Class Intervals for County Income Distributions by Sector, 1950 and 1960

Income class intervals	1950	1960	
	Every sector[a/]	The All family sector	Other family sectors[b/]
	thousand dollars		
1	0- 0.5	0- 1	0- 1
2	0.5- 1.0	1- 2	1- 2
3	1.0- 1.5	2- 3	2- 3
4	1.5- 2.0	3- 4	3- 4
5	2.0- 2.5	4- 5	4- 5
6	2.5- 3.0	5- 6	5- 6
7	3.0- 3.5	6- 7	6- 7
8	3.5- 4.0	7- 8	7- 8
9	4.0- 4.5	8- 9	8- 9
10	4.5- 5.0	9-10	9-10
11	5.0- 6.0	10-15	10+
12	6.0- 7.0	15-25	
13	7.0-10.0	25+	
14	10.0+		

a/ Includes A(f), A(fu), N(f), N(fu), and RF(fu).

b/ Includes N(f) and RF(f).

A = All
N = nonfarm
RF = rural farm

(f) = family
(fu) = family and unrelated

published data only for counties with a rural farm population of 400 or more.[1/] Consequently, the following six counties had no observations in the rural farm sector: Alpine, Inyo, Mono, San Francisco, Sierra, and Trinity.

A number of problems arise in working with Census income data. These include nonresponse (applicable in both phases of the present study) and the estimation of average income given skewed income distributions (applicable in the first phase of this study). Techniques developed to handle these problems are discussed in Appendix A of this report. In addition, Appendix A considers (1) the problem of differences in U. S. Office of Business Economics and Census measures of income and (2) the problem of comparability of income measures based on family income versus measures based on family and unrelated income. The discussion of these problems can be of relevance in applications of results presented here.

Average Income Estimates.--Using the estimating procedures for the Census income distribution data described in Appendix A, average income estimates by county were developed and are presented for 1950 in Table 2 and for 1960 in Table 3. (The county ordering in these tables is on the basis of location, running from north to south.[2/])

Table 2 lists family and unrelated average income for the All, nonfarm, and rural farm sectors; and it lists family average income for the All and rural farm sectors. In Table 3, estimated averages for 1960 All family and rural farm family are presented.

- -

1/ U. S. Bureau of the Census, U. S. Census of Population: 1960 . . ., p. 6-458.

2/ For a map of California counties coded in this order and an alphabetical list of counties and corresponding code numbers, see Figure B-1, infra, p. 91.

TABLE 2

Estimates of Average Income by Area, Unit, and Sector, 1950

Area	Family		Family and unrelated		
	All	Rural farm	All	Nonfarm	Rural farm
	dollars				
California	4,124.000	3,966.000	3,576.000	3,576.000	3,569.000
County					
Modoc	4,032.849	3,948.920	3,376.140	3,410.586	3,291.906
Siskiyou	4,107.168	4,778.150	3,554.400	3,488.027	3,968.584
Del Norte	3,301.635	2,804.535	2,846.743	2,876.193	2,268.547
Humboldt	4,338.955	4,519.689	3,723.578	3,713.562	3,711.657
Trinity	2,605.532	1,797.671	2,274.253	2,372.874	1,562.647
Shasta	3,889.470	3,641.860	3,419.011	3,442.978	3,116.178
Lassen	4,264.895	4,520.818	3,702.610	3,706.122	3,562.642
Plumas	4,253.256	4,538.284	3,637.140	3,621.897	3,763.896
Tehama	3,806.733	4,086.829	3,308.023	3,224.779	3,506.773
Mendocino	3,601.695	3,255.261	3,141.064	3,220.209	2,786.480
Glenn	4,241.340	4,904.962	3,569.225	3,243.206	4,139.224
Butte	3,616.104	3,853.409	3,095.778	3,055.133	3,329.078
Sierra	3,446.267	3,207.753	3,036.939	3,051.326	2,773.618
Nevada	3,205.650	3,087.079	2,743.900	2,745.320	2,639.306
Yuba	3,651.844	3,433.966	3,116.777	3,134.647	2,912.483
Sutter	3,862.822	3,966.115	3,355.406	3,354.272	3,358.981
Colusa	4,602.208	5,613.682	3,683.826	3,447.739	4,346.810
Lake	3,004.350	3,009.783	2,599.997	2,604.656	2,577.578
Sonoma	3,649.113	3,732.304	3,113.712	3,108.837	3,115.170
Napa	3,911.682	4,121.879	3,254.895	3,257.669	3,191.370
Yolo	4,214.408	4,477.933	3,219.018	3,195.497	3,302.013
Placer	3,684.859	3,206.051	3,176.991	3,271.385	2,607.310
El Dorado	3,450.566	3,695.118	2,965.179	2,939.755	3,086.665
Sacramento	4,496.130	3,856.705	3,769.205	3,801.177	2,994.484
Solano	4,017.934	3,831.031	3,331.363	3,362.447	2,759.784
Marin	5,387.828	4,659.208	4,376.271	4,385.935	3,498.760
Contra Costa	4,374.002	5,004.161	3,729.737	3,714.276	3,851.516
San Joaquin	3,948.721	3,727.665	3,257.558	3,322.005	2,863.874
Amador	3,534.891	3,606.326	3,054.226	3,044.208	3,079.731
Calaveras	3,479.342	3,438.014	2,929.172	2,940.365	2,846.782
Alpine	5,640.077	7,661.062	4,605.662	1,999.695	7,633.163
Mono	3,918.629	5,768.445	3,168.218	3,043.793	4,069.529
Tuolumne	3,774.862	4,067.710	3,198.726	3,167.778	3,427.199
Stanislaus	3,576.422	3,818.042	3,225.058	3,169.999	3,441.282
Alameda	4,596.069	4,562.170	3,789.951	3,778.494	3,544.957
San Francisco	4,760.365	a/	3,648.603	3,635.277	
San Mateo	5,568.760	4,256.267	4,958.073	4,952.581	3,320.520

(Continued on next page.)

TABLE 2--continued.

Area	Family		Family and unrelated		
	All	Rural farm	All	Nonfarm	Rural farm
	dollars				
Santa Cruz	3,663.374	3,577.560	3,045.755	3,041.476	2,985.336
Santa Clara	4,389.024	4,280.122	3,587.747	3,586.745	3,463.137
Merced	3,828.666	4,012.537	3,291.835	3,231.871	3,456.058
Mariposa	3,298.972	3,359.639	2,629.923	2,623.858	2,642.868
Madera	3,195.820	3,430.963	2,939.658	2,831.011	3,167.574
Inyo	4,114.026	4,503.324	3,495.759	3,481.523	3,601.101
Fresno	3,835.608	3,681.667	3,414.064	3,442.582	3,249.052
San Benito	3,757.143	3,928.387	3,124.025	3,014.477	3,284.837
Monterey	4,281.370	4,423.380	3,253.688	3,242.762	3,271.210
Kings	3,861.173	3,889.844	3,531.605	3,528.437	3,530.971
Tulare	3,404.748	3,814.849	3,090.419	2,965.625	3,468.637
San Bernardino	3,439.599	3,444.219	3,008.342	3,003.130	2,961.119
Kern	4,159.347	4,084.936	3,690.198	3,696.397	3,516.556
San Luis Obispo	3,608.211	4,595.268	2,954.980	2,846.044	3,728.429
Santa Barbara	4,362.000	4,704.896	3,468.868	3,435.649	3,746.191
Ventura	4,249.011	4,975.050	3,642.771	3,535.299	4,362.227
Los Angeles	4,352.757	4,591.338	3,686.226	3,672.877	3,749.414
Orange	3,857.954	4,915.320	3,347.364	3,276.746	4,271.259
Riverside	3,303.894	3,502.511	2,834.057	2,811.864	2,981.172
Imperial	3,642.966	3,046.302	2,938.354	3,149.866	2,238.778
San Diego	4,050.935	4,590.450	3,212.690	3,192.348	3,478.041

a/ Blanks indicate not applicable.

Source: U. S. Bureau of the Census, U. S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 5, California, 1952, p. 5-71.

TABLE 3

Estimates of Average Family Income by Area
for All and Rural Farm Sectors, 1960

Area	Family	
	All	Rural farm
	dollars	
California	8,000.00	7,295.00
County		
Modoc	6,777.64	8,000.83
Siskiyou	6,391.00	6,859.72
Del Norte	7,316.51	8,975.51
Humboldt	7,407.97	8,109.41
Trinity	6,839.52	a/
Shasta	6,850.89	6,964.57
Lassen	6,524.63	6,882.35
Plumas	6,583.66	6,693.19
Tehama	6,194.71	6,313.83
Mendocino	6,704.68	6,427.63
Glenn	6,252.48	6,485.28
Butte	6,208.15	7,106.60
Sierra	6,161.18	
Nevada	6,253.78	7,179.57
Yuba	5,960.62	6,776.24
Sutter	6,789.29	7,540.84
Colusa	6,939.46	8,134.04
Lake	5,648.90	7,017.75
Sonoma	6,626.77	6,808.85
Napa	7,294.79	8,198.62
Yolo	7,104.44	8,018.47
Placer	6,829.91	6,531.62
El Dorado	7,639.30	7,367.28
Sacramento	8,204.40	8,040.16
Solano	6,838.53	7,072.47
Marin	9,958.70	7,912.54
Contra Costa	8,586.03	8,264.94
San Joaquin	6,903.36	7,383.54
Amador	6,216.77	7,949.03
Calaveras	6,211.82	4,768.74
Alpine	5,143.75	
Mono	7,936.66	
Tuolumne	6,289.88	6,106.78
Stanislaus	6,192.09	6,152.85
Alameda	7,941.92	7,665.73
San Francisco	8,163.15	b/

(Continued on next page.)

TABLE 3--continued.

Area	Family	
	All	Rural farm
	dollars	
San Mateo	9,910.38	10,149.00
Santa Cruz	6,496.60	6,930.65
Santa Clara	8,668.43	7,878.86
Merced	5,997.28	6,172.22
Mariposa	5,474.38	4,999.48
Madera	5,894.79	6,647.11
Inyo	6,241.89	
Fresno	6,746.58	7,009.95
San Benito	6,538.66	7,878.86
Monterey	7,073.64	7,269.34
Kings	6,038.08	6,617.13
Tulare	6,067.08	7,699.88
San Bernardino	6,796.83	7,225.85
Kern	6,901.10	7,375.10
San Luis Obispo	6,578.18	7,555.34
Santa Barbara	8,303.04	8,336.54
Ventura	7,486.35	9,260.36
Los Angeles	8,481.58	8,773.01
Orange	8,409.78	11,229.12
Riverside	6,763.39	7,035.64
Imperial	6,713.11	7,838.28
San Diego	7,663.14	8,266.04

a/ Blanks indicate no data available.

b/ Not applicable.

Source: U. S. Bureau of the Census, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963, p 6-250.

The mean values presented in Tables 2 and 3 are always above the corresponding median values presented in the Census, reflecting the skewness of income distributions. For the All and nonfarm sectors, the ratio of median to mean ranges from about .70 to .90.

For the rural farm sector, however, the ratio ranges from .66 to .86, which can be interpreted as indicating a greater degree of skewness in the farm than in the nonfarm income distribution.

A Technique for One Variable Spatial Analysis Mapping

Given average income by county for a particular sector, a computer program was designed to plot this information on a map of California. The computer program is of general applicability for spatial analysis mapping, though the version employed here applies specifically to counties of the state of California.

In the program, only one variable is analyzed at a time, and discrete classes are formed for otherwise continuous variables. Spatial analysis maps of California are then produced, with a series of maps being produced for each variable of interest. For each variable, a set of maps shows counties which have a high, medium, or low value for that variable. The following criterion was employed in forming these discrete classes. If an observation on a given county is above the mean plus 45 percent of the standard deviation for the sample of all counties, then the given county is assigned to the high classification. Similarly, if the observation is below the mean minus 45 percent of the standard deviation, then the county is assigned to the low classification. If the observation is equal to or between these limits, the county is assigned to the medium classification. The limits selected (by the way of plus or minus 45 percent of the standard deviation) are appropriate in a situation where it is desired to divide a normal distribution into three approximately equal classes. The limits

of the mean plus 45 percent of the standard deviation and the mean minus 45 percent of the standard deviation cut off areas of 32.635 percent of the normal distribution. The middle class then contains approximately 34.73 percent of the area under the normal curve.

The program developed is a general-purpose program and can be used for any variable of interest. In addition, the program is not limited to a three-category classification; as many as nine categories are possible with the present program by changing only one subroutine, and it would be a simple matter to extend the program to handle larger numbers of categories. Again, the mapping can be developed for any state, region, or spatial entity.[1/]

Spatial analysis mappings for 1960 average incomes appear as Figures 1 and 2, with Figure 1 presenting maps for nonfarm income and Figure 2 presenting maps for farm income. The high, medium, and low category counties of the mappings are presented in Tables 4 and 5.

Table 4 lists a cross-classification of counties in terms of nonfarm average income for 1950 and 1960. In the resultant matrix, a county on the main diagonal has remained in the same relative position between periods. This occurs for 11 out of 18 low counties in 1950, 15 out of 25 medium counties in 1950, and 9 out of 15 high counties in 1950. Thus, 35 out of 58 counties are on the main diagonal; that is, they do not change position. Further, only two counties move from low in 1950 to high in 1960, and only three move from high in 1950 to low in 1960. Examining the extremes, it may be noted that the high-income counties are urbanized counties; the low-income counties appear in the central part of the state and generally are clustered around the outer portions of the Central Valley.

- -

1/ For additional references on spatial analysis mapping, see Elliot L. Amidon, A Computer-Oriented System for Assembling and Displaying Land Management Information, U. S. Forest Service Research Paper PSW-17 (Berkeley, California: Pacific Southwest Forest and Range Experiment Station, 1964), 34p.

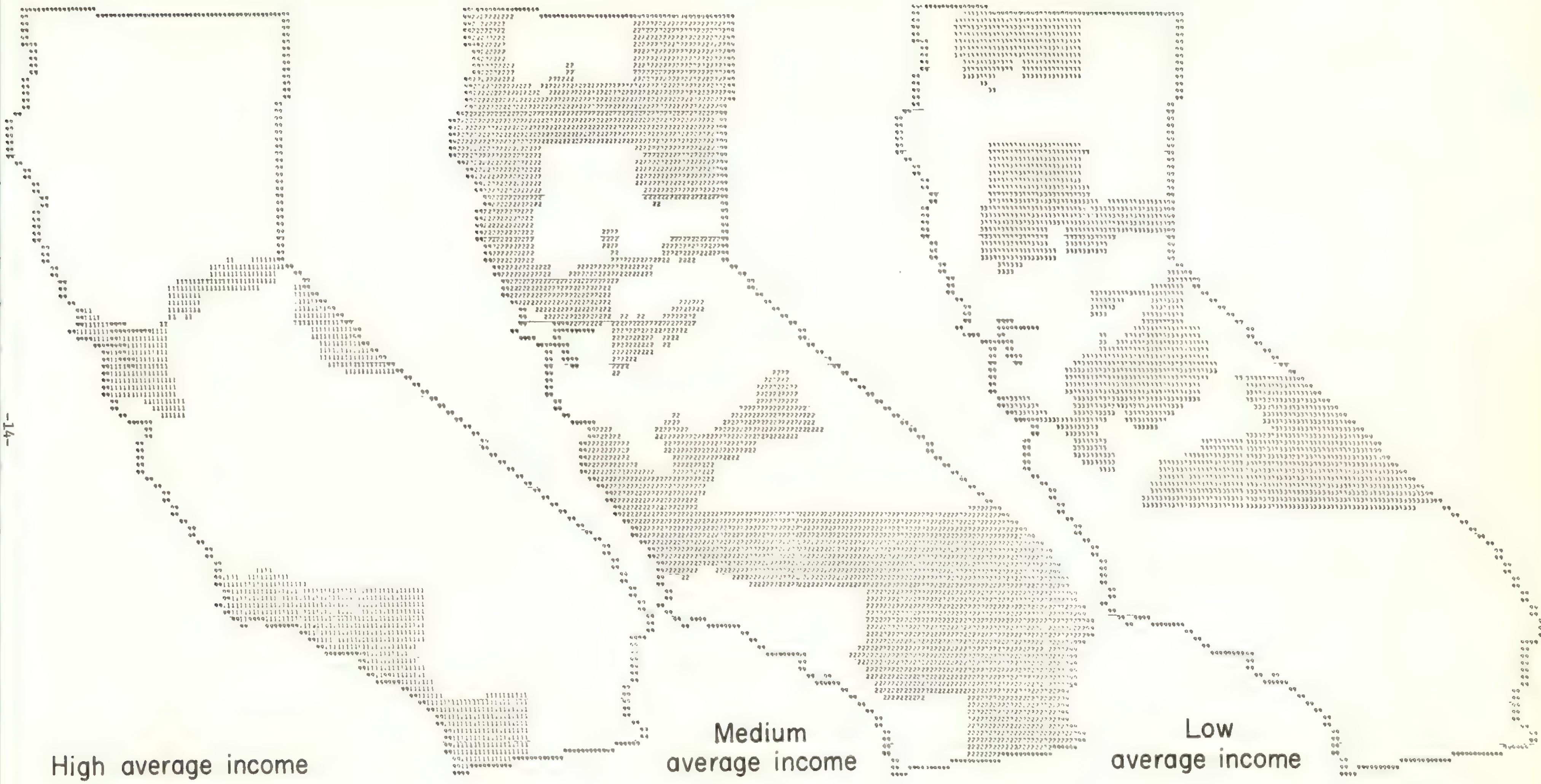


FIGURE 1. Mapping of High, Medium and Low Average Income by County for the Nonfarm Population California, 1960 (Mapping by computer program)

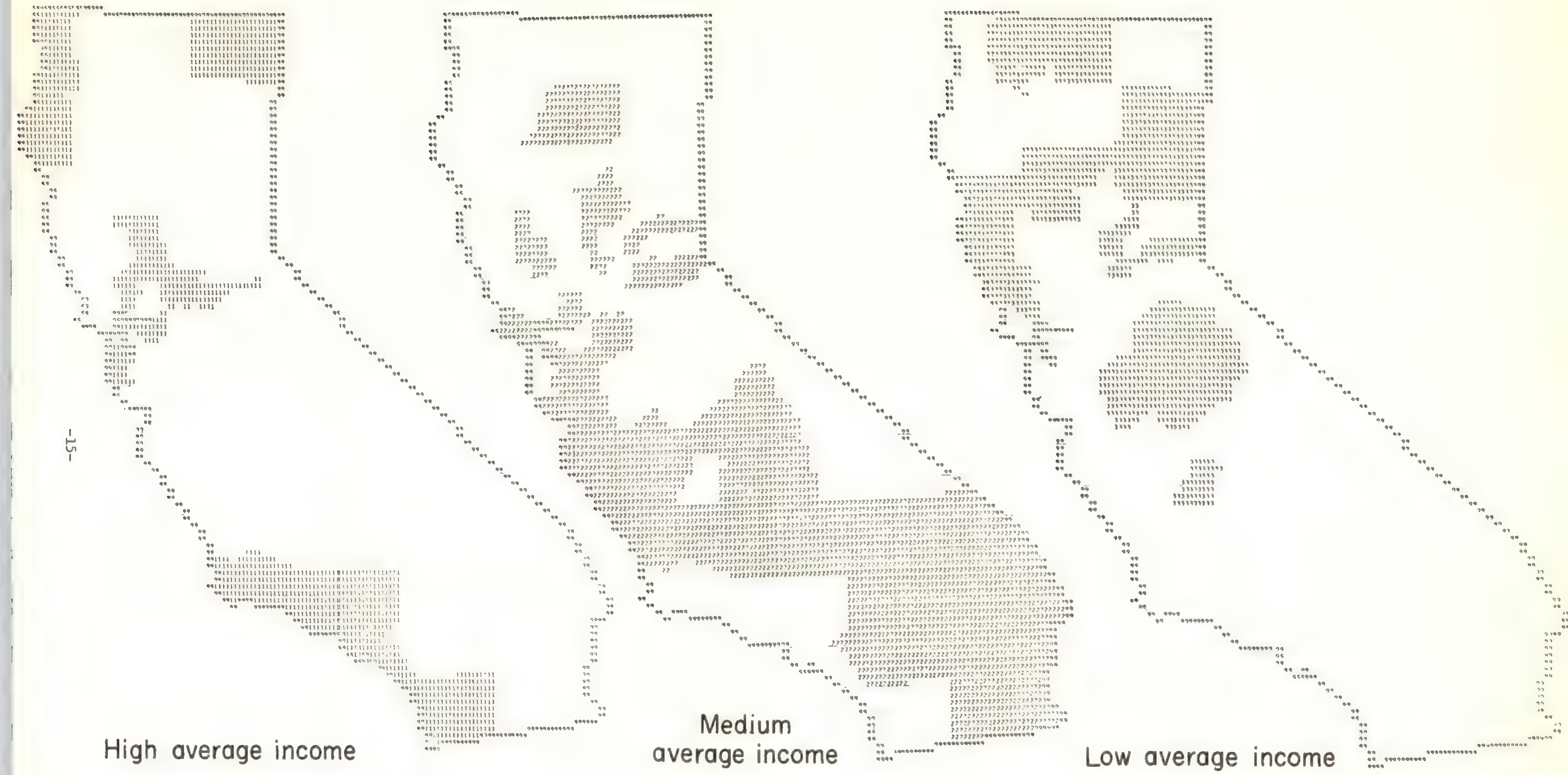


FIGURE 2. Mapping of High, Medium, and Low Average Income by County for the Rural Farm Population California, 1960 (Mapping by computer program)

TABLE 4

Cross-Classification of Rankings of Counties for Average Nonfarm Family Income, 1950 and 1960

Rankings of counties in 1960[a]	Rankings of counties in 1950[a]		
	Low	Medium	High
Low	Butte Sierra Nevada Lake Amador Alpine Mariposa Madera Tulare Santa Cruz San Benito	Tehama Glenn Yuba Stanislaus Merced Tuolumne	Siskiyou Kings Inyo
Medium	Del Norte Trinity Calaveras San Bernardino Riverside	Modoc Shasta Plumas Mendocino Colusa Sutter Placer Yolo Sonoma Solano San Joaquin Fresno Monterey San Luis Obispo Imperial	Humboldt Lassen Kern
High	El Dorado Mono	Napa Santa Barbara Orange San Diego	Sacramento Marin Contra Costa Alameda Santa Clara San Francisco San Mateo Ventura Los Angeles

a/ The number of counties in each category is as follows:

	1950		
1960	Low	Medium	High
Low	11	6	3
Medium	5	15	3
High	2	4	9

Sources: Developed from data in U. S. Bureau of the Census, U. S. Census of Population: 1950, Vol II, Characteristics of the Population, Part 5, California, 1952, p. 5-71; and idem, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963, p. 6-250.

TABLE 5

Cross-Classification of Rankings of Counties for Average Rural Farm Family Income, 1950 and 1960

Rankings of counties in 1960a/	Rankings of counties in 1950a/		
	Low	Medium	High
Low	Mendocino Yuba Placer Calaveras Mariposa	Tehama Lassen Sonoma Stanislaus Tuolumne Merced Madera Kings	Siskiyou Plumas Glenn
Medium	Lake Nevada Solano San Joaquin Santa Cruz San Bernardino Riverside Imperial	Shasta Butte Sutter Marin El Dorado San Benito Alameda Santa Clara Fresno Monterey Tulare Kern	San Luis Obispo
High	Del Norte Sacramento	Modoc Yolo Napa Amador San Mateo San Diego	Humboldt Colusa Contra Costa Santa Barbara Ventura Los Angeles Orange
b/	Trinity Sierra	Inyo	Alpine Mono

a/ San Francisco County is excluded, having no rural farm sector. The number of counties in each category is as follows:

1960	1950 Low	1950 Medium	1950 High
Low	5	8	3
Medium	8	12	1
High	2	6	7
No data	2	1	2

b/ No 1960 data available for these counties.

Sources: Developed from data in U. S. Bureau of the Census, U. S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 5, California, 1952, p. 5-71; and idem, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963, p. 6-250.

Table 5 lists a cross-classification of counties in terms of farm average income for 1950 and 1960. There is again some evidence of regional clustering, but farm income appears much more variable than nonfarm income. For counties with data available, 5 out of 15 low counties, 12 out of 26 medium counties, and 7 out of 11 high counties for 1950 remain in the same classification in 1960. Thus, only 24 out of 52 counties are on the main diagonal; that is, they do not change position. As a fraction, this is .46 as compared to the nonfarm fraction of .60.

Southern California coast counties are in the high category as are Contra Costa, Colusa, and Humboldt. The persistently low-income counties include Mendocino and four counties on the eastern edge of the Central Valley.

The mappings tend to confirm the T. W. Schultz location hypothesis, which involves the argument that average farm income is a function of proximity to metropolitan markets, that income increases as access increases.[1/] This relationship seems particularly strong for the 1960 case.

Intersectorial and Intertemporal Comparisons

Given the results developed in the preceding section, a number of comparisons between sectors and over time became possible and were carried out.

Intersectorial comparisons were carried out through comparison of rural farm to nonfarm results for both 1950 and 1960 data. Ratios of average income for farm relative to nonfarm are presented in Table 6.[2/] (It should be noted

- -

1/ Theodore W. Schultz, The Economic Organization of Agriculture (New York: McGraw-Hill Book Company, Inc., 1953), Chaps. IX, X, XVII, and XVIII.

2/ Computation of the necessary mean values employed here involved some additional data manipulation. In particular, some income classes were aggregated so as to make use of identical average values for the upper tail of the empirical distributions. For this reason, the means used here are not necessarily identical with those presented earlier in this section.

TABLE 6

Intersector Ratio Comparisons of Average Family Income:
Rural Farm Relative to Nonfarm Sector by County
1950 and 1960

County	Family and unrelated, 1950	Family, 1960
	Ratio of rural farm to nonfarm[a]	Ratio of rural farm to nonfarm[a]
Modoc	0.96	1.13
Siskiyou	1.13	1.00
Del Norte	0.78	1.18
Humboldt	0.99	1.02
Trinity	0.65	b/
Shasta	0.90	0.93
Lassen	0.96	0.95
Plumas	1.03	0.92
Tehama	1.08	0.94
Mendocino	0.86	0.89
Glenn	1.27	1.00
Butte	1.09	1.08
Sierra	0.90	
Nevada	0.96	1.07
Yuba	0.92	1.05
Sutter	1.00	1.04
Colusa	1.26	1.17
Lake	0.99	1.23
Sonoma	1.00	0.95
Napa	0.98	1.03
Yolo	1.03	1.02
Placer	0.79	0.86
El Dorado	1.05	0.87
Sacramento	0.78	0.87
Solano	0.82	0.93
Marin	0.79	0.73
Contra Costa	1.03	0.87
San Joaquin	0.86	1.00
Amador	1.01	1.15
Calaveras	0.96	0.67
Alpine	3.81	
Mono	1.33	
Tuolumne	1.08	0.88
Stanislaus	1.08	0.93
Alameda	0.93	0.88
San Francisco	c/	c/
San Mateo	0.67	0.94

(Continued on next page.)

TABLE 6--continued.

County	Family and unrelated, 1950 Ratio of rural farm to nonfarm[a]	Family, 1960 Ratio of rural farm to nonfarm[a]
Santa Cruz	0.98	1.01
Santa Clara	0.96	0.81
Merced	1.06	0.97
Mariposa	1.00	0.90
Madera	1.11	1.12
Inyo	1.03	
Fresno	0.94	0.97
San Benito	1.09	1.18
Monterey	1.00	0.97
Kings	1.00	1.04
Tulare	1.17	1.27
San Bernardino	0.98	0.96
Kern	0.95	0.98
San Luis Obispo	1.31	1.08
Santa Barbara	1.09	0.93
Ventura	1.23	1.13
Los Angeles	1.02	0.95
Orange	1.30	1.21
Riverside	1.06	0.96
Imperial	0.71	1.10
San Diego	1.08	0.98

a/ Ratio of rural farm to (urban plus rural nonfarm).

b/ Blanks indicate no data reported for the counties involved.

c/ Not applicable.

that the 1950 results are based on family and unrelated income, while those for 1960 are based on family income. Because of differing proportions of family and of unrelated units in farm and nonfarm populations, these bases are not strictly equivalent. Use of the family and unrelated base will tend to raise the farm average relative to the nonfarm. This qualification should be kept in mind in considering the results below.) The ratio comparison for 1960 is exhibited in the map of Figure 3.

On the whole, California rural farm and nonfarm average incomes do not diverge particularly. Thus, for 1950, 31 counties have an income ratio above 1.0 and 26 counties have an income ratio below 1.0; in 1960, the distribution is 24 and 28, respectively. Again, for 1950, 37 counties (out of 57) have an income ratio between .9 and 1.1; for 1960, 32 counties (out of 52) fall within these limits. There are four counties with a ratio below .9 in both years (Mendocino, Placer, Sacramento, and Marin) and five counties with a ratio above 1.1 in both years (Colusa, Madera, Tulare, Ventura, and Orange).

Growth in average income over time was measured and compared using the compound growth rate per year. The compound growth rate was estimated for each county using the relation:

$$\overline{Y}_{'60} = (1 + r)^{10}\ \overline{Y}_{'50}$$

where

$\overline{Y}_t$ = average income

t = 1950 and 1960

and

$(1 + r)$ = yearly compounding factor.

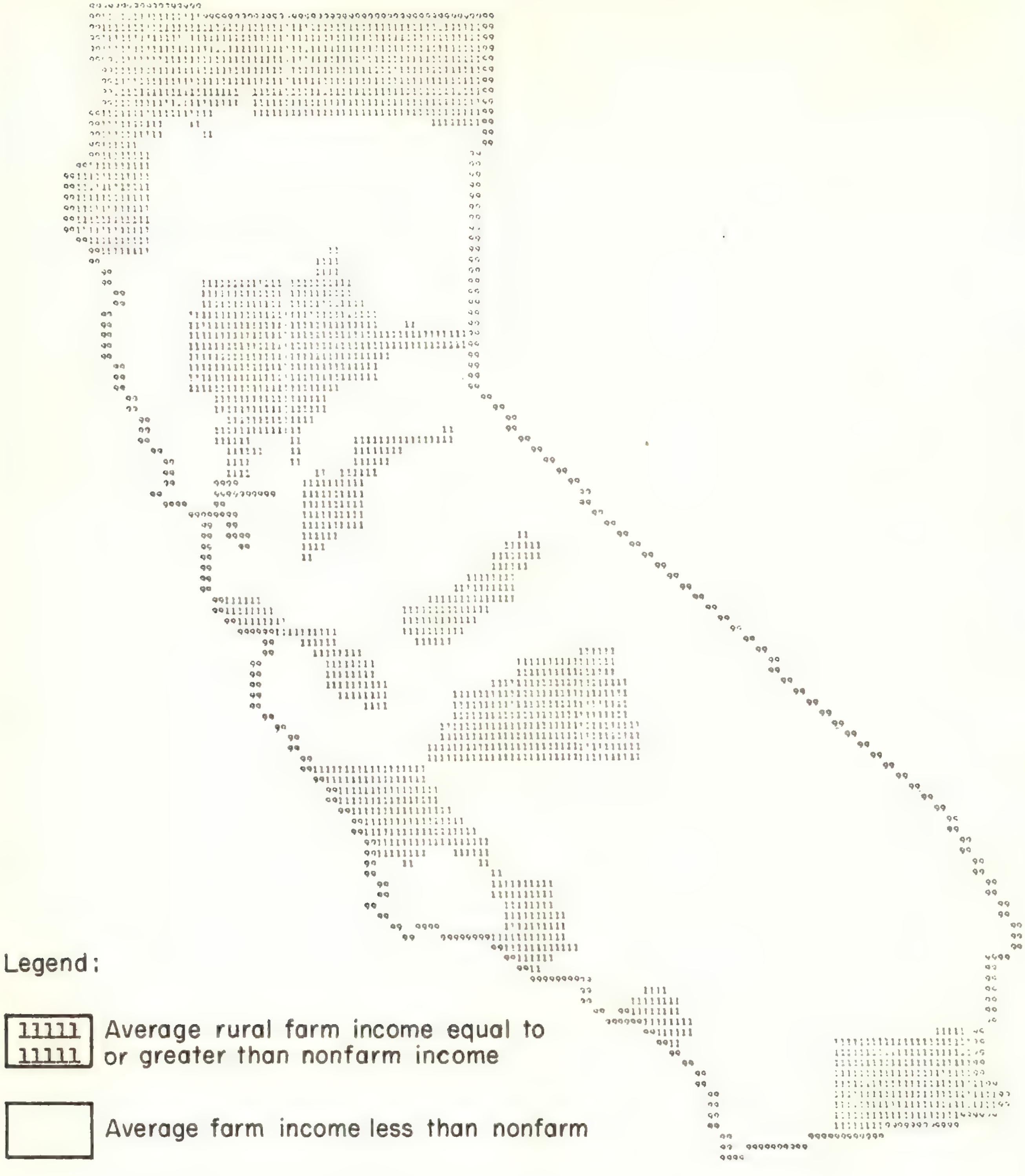


FIGURE 3. Comparison of Rural Farm to Nonfarm Average Income for the Family Observation Unit by County, 1960
(Mapping by computer program)

Results obtained appear in Table 7 for the nonfarm and the rural farm sectors using family income estimates.

It is worth noting that the data in Table 7 are not deflated for changes in the general price level. Such deflation can be carried out with relative ease if estimates of growth in real income are desired.[1/]

For rural farm income, counties with high rates of growth appear to fall into regional clusters. Thus, counties with growth rates above 7 percent (in Table 7) include clusters in the southeastern part of the state, in the northern San Joaquin area, and above the Bay area (Napa, Lake, and Mendocino counties).

If rural farm growth rates are compared to nonfarm growth rates, it turns out that the rural farm rate is higher for 22 counties and lower for 30 counties (of the 52 counties for which data are available). Regional patterns seem fairly well defined, with farm rates above nonfarm for the southern San Joaquin Valley, the eastern part of the Sacramento Valley, and for some scattered counties at the extreme portions of the state.

A final set of comparisons can be obtained by examining changes in average income rank between 1950 and 1960. (Income rates were obtained from the data of Tables 2 and 3. They are presented and discussed in Appendix B.) Comparisons were carried out by subtracting the 1960 rank from the 1950 rank; if a county moved up in rank (and, hence, had a lower rank number), this difference is positive. Similarly, a decline in rank yields a negative difference. Results are plotted in Figures 4 and 5.

- -

1/ For example, on a 1957-1959 base, the Bureau of Labor Statistics consumer price index increased from 83.0 to 101.5 between 1949 and 1959. These figures yield a compound growth rate in the price level, p, equal to 2.03 percent per year. The deflated income growth rate, r^*, is then found from the following equation: $(1 + r) = (1 + r^*)(1 + p)$.

TABLE 7

Estimated Yearly Growth Rates of Average Family Income for Rural Farm and Nonfarm Sectors by County for the Period 1950-1960

County	Compound growth rate, r	
	Nonfarm	Rural farm
Modoc	5.66	7.32
Siskiyou	5.52	3.68
Del Norte	8.63	12.34
Humboldt	6.27	6.02
Trinity	11.06	a/
Shasta	6.66	6.70
Lassen	5.48	4.29
Plumas	5.54	3.96
Tehama	6.09	4.45
Mendocino	6.99	7.04
Glenn	5.33	2.83
Butte	6.30	6.31
Sierra	7.13	
Nevada	7.63	8.81
Yuba	5.77	7.03
Sutter	6.60	6.64
Colusa	4.98	3.78
Lake	6.60	8.83
Sonoma	7.04	6.20
Napa	7.47	7.12
Yolo	6.52	6.00
Placer	7.28	7.38
El Dorado	9.44	7.14
Sacramento	7.36	7.62
Solano	6.56	6.32
Marin	7.14	5.44
Contra Costa	8.13	5.15
San Joaquin	6.36	7.07
Amador	6.95	8.22

(Continued on next page.)

TABLE 7--continued.

County	Compound growth rate, r	
	Nonfarm	Rural farm
Calaveras	7.32	3.33
Alpine	5.98	
Mono	8.64	
Tuolumne	6.29	4.15
Stanislaus	6.46	4.89
Alameda	6.56	5.33
San Francisco	6.34	b/
San Mateo	6.79	9.08
Santa Cruz	6.48	6.84
Santa Clara	8.20	6.29
Merced	5.37	4.40
Mariposa	5.42	4.06
Madera	6.73	6.84
Inyo	5.02	
Fresno	6.40	6.65
San Benito	6.22	7.21
Monterey	5.77	5.09
Kings	5.14	5.46
Tulare	6.38	7.28
San Bernardino	8.10	7.69
Kern	6.07	6.09
San Luis Obispo	7.23	5.10
Santa Barbara	7.54	5.89
Ventura	6.99	6.41
Los Angeles	7.78	6.69
Orange	9.38	8.61
Riverside	8.33	7.22
Imperial	6.46	9.91
San Diego	7.60	6.06

a/ Blanks indicate no data available.

b/ Not applicable.

Sources: Developed from data in U. S. Bureau of the Census, U. S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 5, California, 1952; and idem, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963.

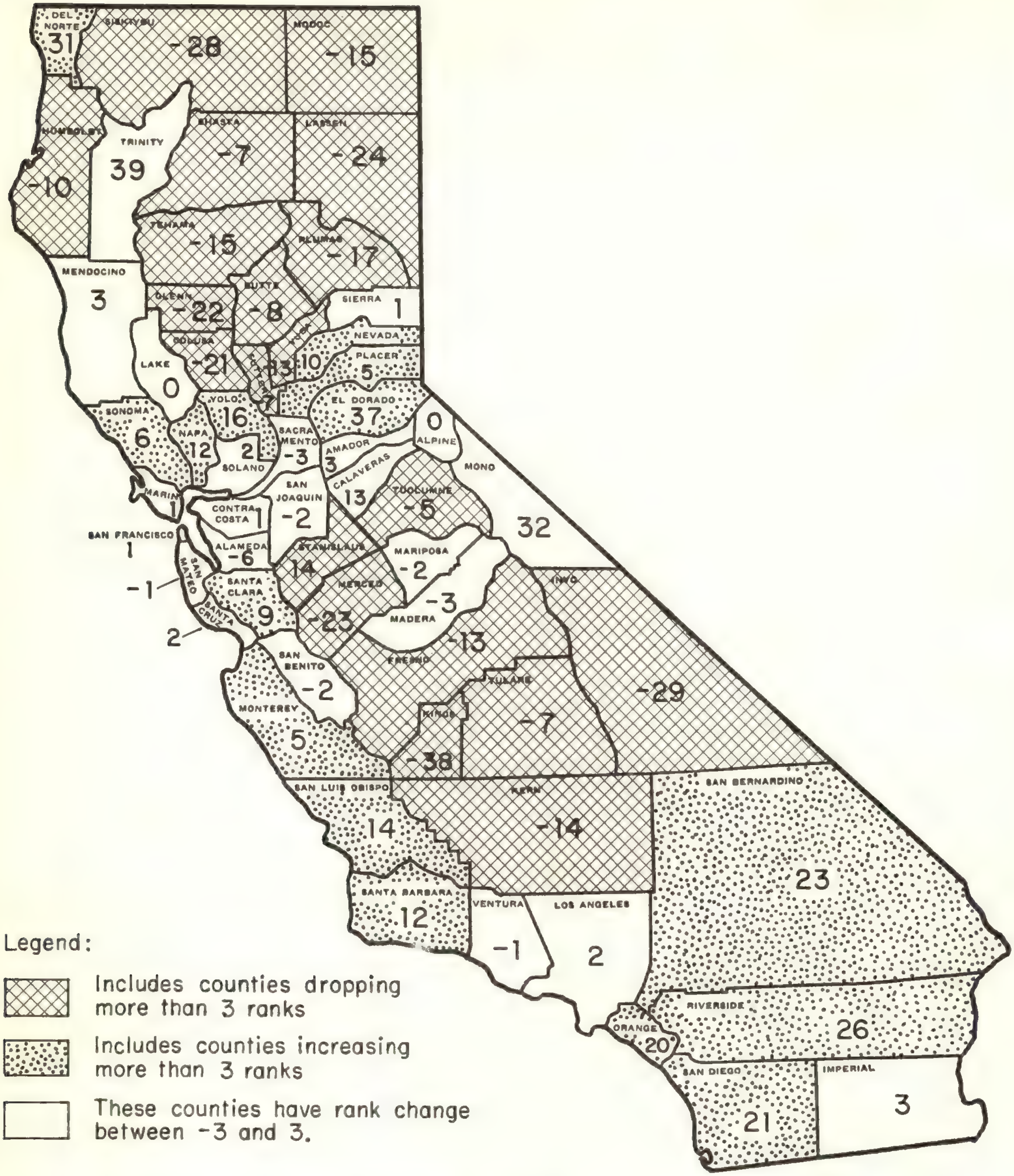


FIGURE 4 Change in Rank of Counties for Nonfarm Income between 1950 and 1960 (numbers indicate amount of change in rank; thus, a positive number indicates a county is higher in rank ... relatively better off)

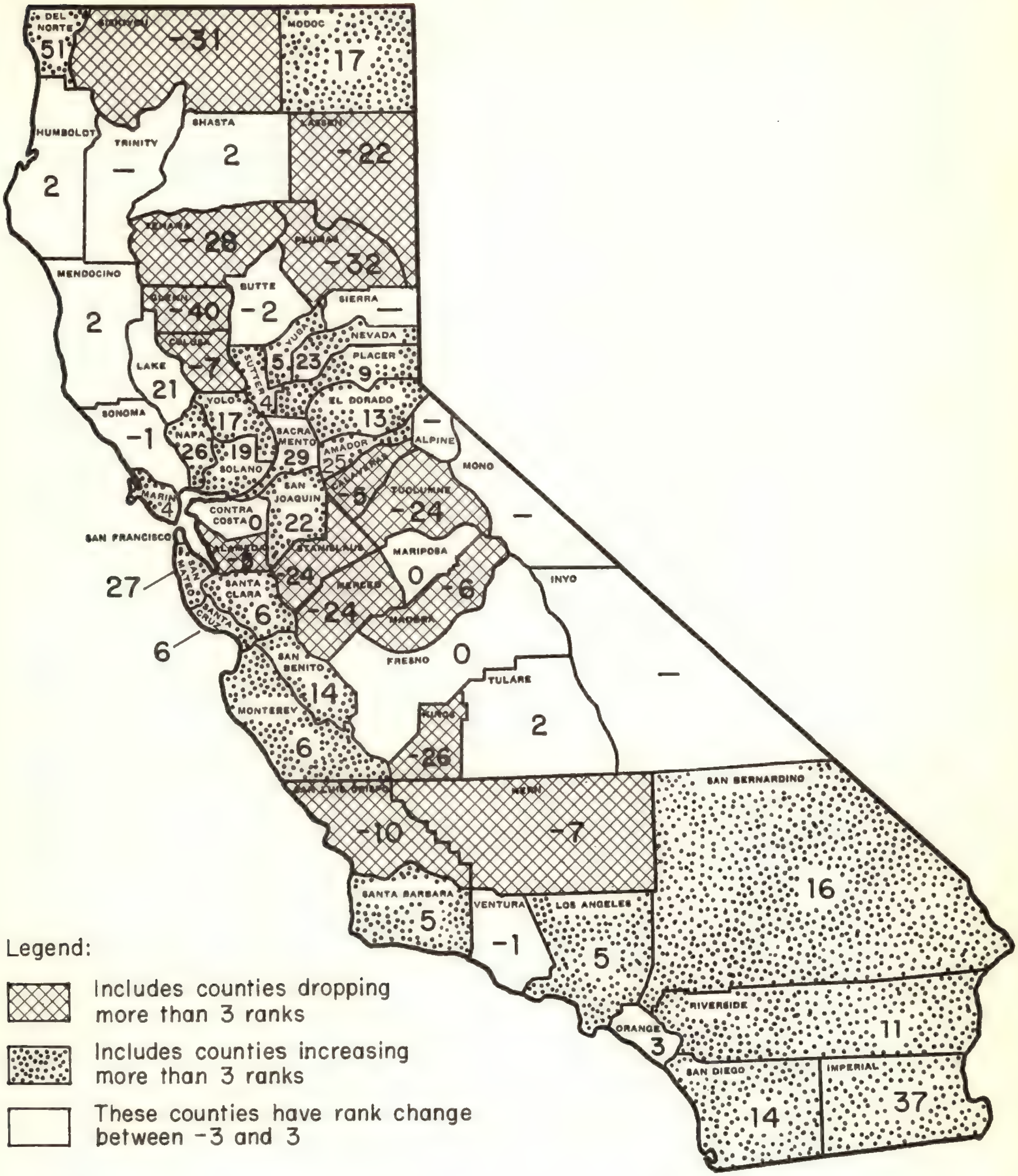


FIGURE 5 Change in Rank of Counties for Rural Farm Income between 1950 and 1960 (numbers indicate amount of change in rank; thus, a positive number indicates a county is higher in rank ··· relatively better off)

The results can be interpreted as indicating changes in relative position. There seem to be well-defined regional patterns in Figures 4 and 5.

Thus, the farm and nonfarm patterns show marked similarity. Increased rank (higher relative position) occurs for southern California (particularly in proximity to Los Angeles), for some counties in the Bay area, and for counties near Sacramento County. Decreased rank occurs in the San Joaquin Valley and in the northern part of the state. A fairly obvious hypothesis here is that increased urbanization (and suburbanization) is the major factor at work in these shifts. It is worth stressing that these shifts occur for the farm as well as nonfarm sectors.

The work of this section concludes with a brief investigation of the hypotheses that (1) growth in average family income is related to base period (1950) income and (2) level of rural farm average income is related to nonfarm average income.

It was hypothesized that growth in average income is negatively related to base period income, reflecting equilibrating forces at work in the economy. The hypothesis was tested by regression of income growth from 1950 to 1960 on base year income (1950) by county. This regression equation was fitted for both the nonfarm and the rural farm sectors. Regression results obtained appear in Table 8.

The results obtained were in line with expectations in that a negative slope was obtained in each equation. The R^2 values were not particularly high; however, t-ratios were significant at the 5-percent level for the rural farm sector and close to significant for the nonfarm sector.

The hypothesis that level of rural farm average income is related to nonfarm average income was also tested by regression analysis. It was expected that slopes would be positive and this in fact occurred. Increasing levels of nonfarm income <u>are</u> associated with increasing levels of farm income.

TABLE 8

Regression Results Obtained in Relating County Growth Rate of Average Income (for the Period 1950-1960) to Base Year Average Income (1950) by Nonfarm and Rural Farm Sectors

Sectors	Number of county observations[a]	R^2	Constant	Coefficient	t-ratio
Nonfarm	58	.052	8.767	-.00050	-1.750
Rural farm	52	.307	12.871	-.00162	-4.708

a/ Six counties had no data on rural farm income in 1960.

Table 9 lists the results obtained for rural farm income regressed on nonfarm income for both 1950 and 1960. It also lists results obtained when rural farm growth rates are regressed on nonfarm growth rates. In all cases, estimated coefficients are significantly different from zero at the 5-percent level.

A Multivariate Income Distribution Determination Model

The second major phase of this study is concerned with the construction and application of a multivariate model of factors determining the size distribution of income. The empirical application here involves the California rural farm sector for 1950 and 1960 (using family income data), but it is expected the model will be of general interest and applicability.

Material covered in this section is organized into the following parts: (1) a discussion of previous theoretical and empirical work and a presentation of the general approach used here, (2) a formal summary of the statistical model and of tests of hypotheses for that model, (3) a discussion of the specific form of the applied model and a statement of results obtained, and (4) interpretations

TABLE 9

Regression Results Obtained in Relating Rural Farm Average Income Measure to Nonfarm Average Income Measure, 1950 and 1960

Period and income measure a/	Number of county observations b/	R^2	Constant	Coefficient	t-ratio
1950 average income	52	.366	1,152	.73762	5.38
1960 average income	52	.171	4,178	.40222	3.21
Annual rate of growth in average income for the period 1950-1960	52	.270	0.2445	.90378	4.30

a/ The measure listed is used in the regression of rural farm values on nonfarm values; the rural farm measure is the dependent variable, and the nonfarm measure is the independent variable.

b/ Six counties had no data available on rural farm income in 1960; these counties were also excluded from the 1950 regression for consistency in comparison.

of those results. This study then concludes with a summary of significant findings and some suggestions on the direction of future work.

Size Distribution of Income: Previous Work and Approach Used Here

No comprehensive theory of the size distribution of income exists at the present date. Even though attempts at theorizing began as early as the Physiocrats, no complete theory has been developed.[1/] Two approaches have emerged over the years. One is concerned with the functional distribution of income, while the other is concerned with the personal (or size) distribution of income. Functional distribution of income theories are the best known for they center around economic theory and the marginal pricing of the factors of production. However, the case is usually abstracted to a single income-earning unit with the objective of explaining that unit's level of income in terms of conventional theory. The other area of theoretical work, which has concentrated on explaining the shape of the personal distribution of income, is relatively in its infancy.

An example of theoretical work in this area is that of Robert Summers, who was concerned with estimating the shape of lifetime size distributions and isolating some of the relevant factors in determining that shape.[2/] Thus, the problem was to construct a theoretical framework which could be used to estimate the prospective lifetime size distribution of average annual income for a "household" which remained the same in an economy that did not change. The approach is temporally dynamic but assumes no structural changes in the economy.

- -

1/ For a historical perspective, see John Maurice Clark, "Distribution," Readings in the Theory of Income Distribution ("Blakiston Series of Republished Articles on Economics," Vol. III; Philadelphia, Pennsylvania: The Blakiston Company, 1949), Chap. 3, pp. 58-71.

2/ Robert Summers, An Econometric Investigation of the Size Distribution of Lifetime Average Annual Income, Stanford University, Department of Economics, Technical Report No. 31 (Stanford, California, 1956), 127p.

Empirical work in the area has involved attempts at direct analyses of the influence of selected variables upon certain income statistics or indirect analyses by first standardizing the distribution, recomputing the statistic from the distribution, and then examining the effect upon the statistic. An example of the first approach is the work of Zeman.[1/] Zeman made adjustments for certain influencing variables directly upon his indices of wages. An example of the second approach is that of Smolensky, who first standardized the income distribution and then examined the effect upon a statistic of the distribution (the coefficient of variation).[2/]

The above general approaches have the disadvantage of not indicating or measuring the functional relation between the supposedly influencing variables and the income distribution. Thus, the statistical significance of the findings cannot be tested.

F. Gerard Adams moved toward statistical tests of functional relations involving income distributions by constructing a model which "explains" the "amount of an individual's income."[3/] This model can be used to make statistical tests of relations between explanatory variables and the variable amount of individual income. But it cannot be used to make statistical tests of the influence of explanatory variables upon the distribution of income.

- -

1/ Morton Zeman, "A Quantitative Analysis of White-Nonwhite Income Differentials in the United States" (unpublished Ph.D. dissertation, Department of Economics, University of Chicago, 1955). A discussion of Zeman's dissertation appears in Robert J. Wolfson, "An Econometric Investigation of Regional Differentials in American Agricultural Wages," Econometrica, Vol. 26, No. 2 (April, 1958), pp. 225-257.

2/ Eugene Smolensky, "An Interrelationship Among Income Distributions," Review of Economics and Statistics, Vol. XLV, No. 2 (May, 1963), pp. 197-206.

3/ F. Gerard Adams, "The Size of Individual Incomes--Socio-Economic Variables and Chance Variation," Review of Economics and Statistics, Vol. XL, No. 4 (November, 1958), pp. 390-398.

Work here is concerned with estimating functional relations and in developing statistical tests of the influence of explanatory variables on income distributions. It differs from previous empirical work by attempting to relate observations in a given income interval to a set of explanatory variables. This would appear to be an innovation in income distribution studies, and some details of the procedure are worth noting at this point.

Census data on family unit income distributions by county for the rural farm sector were employed in the first section of this study to construct county averages. They are employed again in this section as observations on a set of dependent variables. Each income class interval is associated with a corresponding dependent variable. The dependent variable consists of the proportionate number or frequency of income units measured in percentage terms. For each dependent variable, then, there will be as many observations as counties reported.

Each of the dependent variables is related to a given set of independent variables using multivariate analysis. Observations on these variables are the same for all cases; that is, each dependent variable is related to the same set of independent variables.

The meaning of coefficients obtained is of some interest. A negative coefficient indicates that, as the independent variable increases, the number of units falling in the specific income interval decreases. Similarly, a positive coefficient indicates an increasing number of units in the interval as the independent variable increases. The multivariate approach employed here allows an investigation of the behavior of a coefficient in each income class in turn. A good deal of additional information (relative to work with income class averages) is thereby generated. With results obtained for different time periods, changes in structure can then be investigated.

The Statistical Model

The statistical model and hypothesis-testing approach of this section is an application of the theoretical and empirical works of Anderson, Goldberger, Hooper and Zellner, Hotelling, Kendall, O'Regan, Schatzoff, and Wilks.[1/]

Consider the following general linear model formulation for any sector. For G income intervals, we define the following set of G equations:

$$
\begin{aligned}
Y_{(1)} &= \beta_{(1)0} + \beta_{(1)1} X_1 + \beta_{(1)2} X_2 + \cdots + \beta_{(1)K} X_K + u_{(1)} \\
Y_{(2)} &= \beta_{(2)0} + \beta_{(2)1} X_1 + \beta_{(2)2} X_2 + \cdots + \beta_{(2)K} X_K + u_{(2)} \\
Y_{(3)} &= \beta_{(3)0} + \beta_{(3)1} X_1 + \beta_{(3)2} X_2 + \cdots + \beta_{(3)K} X_K + u_{(3)} \\
&\vdots \\
Y_{(G)} &= \beta_{(G)0} + \beta_{(G)1} X_1 + \beta_{(G)2} X_2 + \cdots + \beta_{(G)K} X_K + u_{(G)}
\end{aligned}
$$

1/ T. W. Anderson, An Introduction to Multivariate Statistical Analysis (New York: John Wiley and Sons, Inc., 1958), 374p.

Arthur S. Goldberger, Econometric Theory (New York: John Wiley and Sons, Inc., 1964), 399p.

J. W. Hooper and Arnold Zellner, "The Error of Forecast for Multivariate Regression Models," Econometrica, Vol. 29, No. 4 (October, 1961), pp. 544-555.

Harold Hotelling, "A Generalized T Test and Measure of Multivariate Dispersion," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (Berkeley and Los Angeles: University of California Press, 1951), pp. 23-42.

Maurice George Kendall, A Course in Multivariate Analysis (London: Charles Griffin and Company, Limited, 1957), 185p.

William Gerard O'Regan, "An Experimental Approach to the Determination of Demand for Orange Concentrate" (unpublished Ph.D. dissertation, Department of Agricultural Economics, University of California, Berkeley, 1962), 74p.

Martin Schatzoff, "Sensitivity Comparisons Among Tests of the General Linear Hypothesis," Journal of the American Statistical Association, Vol. 61, Part 1 (June, 1966), pp. 415-435.

Samuel Stanley Wilks, "Certain Generalizations in the Analysis of Variance," Biometrika, Vol. 24 (November, 1932), pp. 471-494.

where a subscript in parentheses denotes income interval, g; g = 1, ..., G. (Parentheses enclose g values for purposes of clarification.)

This set of G equations can be written compactly as:

$$Y = X\beta + U$$

where Y is an N by G regressand observation matrix (N being the number of observations):

$$Y = \left\{Y_{ig}\right\} = \begin{bmatrix} Y_{1(1)} & Y_{1(2)} & \cdots & Y_{1(G)} \\ Y_{2(1)} & Y_{2(2)} & \cdots & Y_{2(G)} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ Y_{N(1)} & Y_{N(2)} & \cdots & Y_{N(G)} \end{bmatrix} \qquad \begin{matrix} i = 1, \ldots, N \\ g = 1, \ldots, G \\ Y_{NXG}. \end{matrix}$$

Thus, the first column consists of all observations in the first income interval. In the present case, this consists of the 52 county observations falling in the lowest income class. A given row, i, then contains all the income interval observations for a given county, i, with i = 1, ..., 52 (given observations on 52 counties). Thus, the first column in the definition of Y corresponds to the set of observations on $Y_{(1)}$ in the first equation of the linear model, and so on through the G<u>th</u> column in the definition of Y. Income interval subscripts are again enclosed in parentheses.

In matrix terms, X is an N by K + 1 regressor observation matrix, with observations on X <u>invariant</u> over income class:

$$X = \left\{ X_{ik} \right\} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1K} \\ 1 & X_{21} & \cdots & X_{2K} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 1 & X_{N1} & \cdots & X_{NK} \end{bmatrix} \qquad \begin{matrix} i = 1, \ldots, N \\ k = 1, \ldots, K. \end{matrix}$$

The coefficient matrix, β, is of size K + 1 by G:

$$\beta = \left\{ \beta_{gk} \right\} = \begin{bmatrix} \beta_{(1)0} & \beta_{(2)0} & \cdots & \beta_{(G)0} \\ \beta_{(1)1} & \beta_{(2)1} & \cdots & \beta_{(G)1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \beta_{(1)K} & \beta_{(2)K} & \cdots & \beta_{(G)K} \end{bmatrix} \qquad \begin{matrix} g = 1, \ldots, G \\ k = 1, \ldots, K. \end{matrix}$$

The disturbance terms appear in an N by G residual matrix:

$$U = \left\{ u_{ig} \right\} = \begin{bmatrix} u_{1(1)} & \cdots & u_{1(G)} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ u_{N(1)} & \cdots & u_{N(G)} \end{bmatrix} .$$

The following assumptions are made with regard to the disturbance terms u_{ig}:

$$E\,(u_{ig}) = 0.$$

For fixed g and i varying, disturbances are assumed distributed independently with equal variance σ_{gg}. Thus,

$$E\left\{u_{ig}\, u_{ig}'\right\} = E \begin{bmatrix} u_{1(g)} \\ \cdot \\ \cdot \\ \cdot \\ u_{N(g)} \end{bmatrix} [u_{1(g)} \cdots u_{N(g)}]$$

$$= \begin{bmatrix} \sigma_{gg} & 0 & \cdots & 0 \\ 0 & \sigma_{gg} & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \sigma_{gg} \end{bmatrix} = \sigma_{gg} I.$$

In similar fashion, for specific county values, g and h:

$$E\left\{u_{ig}\, u_{ih}'\right\} = E \begin{bmatrix} u_{1(g)} \\ \cdot \\ \cdot \\ \cdot \\ u_{N(g)} \end{bmatrix} [u_{1(h)} \cdots u_{N(h)}] = \sigma_{gh} I.$$

Thus, for the same observation number, disturbances in different equations have covariance σ_{gh}; for different observation numbers, disturbances in different equations are independent.

For any observation number, say $i = k$, we may write:

$$E\left\{u_{k(g)}\, u_{k(h)}'\right\} = E \begin{bmatrix} u_{k(1)} \\ \cdot \\ \cdot \\ \cdot \\ u_{k(G)} \end{bmatrix} [u_{k(1)} \cdots u_{k(G)}]$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1G} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \sigma_{G1} & \sigma_{G2} & \cdots & \sigma_{GG} \end{bmatrix} \equiv \left\{\sigma_{gh}\right\} = \Sigma_{GXG}.$$

The general multivariate case does <u>not</u> assume $\sigma_{gh} = 0$ for $g \neq h$; if that assumption were made, the case resulting is that of a set of ordinary least-squares regression equations. If the disturbances u_{ig} are assumed normal, the multivariate case becomes the classical multivariate normal linear regression model.

The β matrix may be estimated by the maximum likelihood method.[1/] The final solution of this mathematical derivation is:

$$X'X\hat{\beta} = X'Y$$

where $\hat{\beta}$ is the matrix of estimated coefficients which, of course, corresponds to the least-squares estimated coefficient matrix. Goldberger develops the proof that individually minimizing the diagonal elements of $\ddagger$ corresponds to minimization of the determinant of $\ddagger$.[2/]

The estimated variance-covariance matrix may be defined as follows: Let $\hat{Y} = X\hat{\beta}$ and $V = Y - \hat{Y}$, then $\hat{\ddagger} = V'V$/degrees of freedom.

In computing results a stepwise procedure was employed. In this procedure the "best" explanatory variable is introduced at the first step, the "second best" explanatory variable at the second step, and so on. The criterion employed in selection of the variable to be introduced was that of minimization of the determinant of the estimated variance-covariance matrix, $|\hat{\ddagger}|$. It should be noted that the employment of this criterion poses a problem of singularity, owing to the definition of Y in terms of frequencies.[3/]

- -

1/ For a recent presentation of this derivation, see Goldberger, *op. cit*.

2/ *Idem*, pp. 205 and 206.

3/ In the Y matrix for this case, each row sums to 100 and each column sums to 100. This results in a $(Y - \overline{Y})$ matrix with the characteristics that (1) each row sums to zero and (2) each column sums to zero. When these characteristics hold for the basic matrix, the resulting matrix of sums of squares and cross products will have the same characteristics. As a consequence, any row will exactly equal the negative sum of all other rows.

To proceed with estimation in this situation, it is only necessary to eliminate one column of the Y matrix. Estimation can then be carried out for a set of G - 1 equations. To obtain the vector of β coefficients omitted in the first phase, an alternative column is eliminated and estimates obtained for the alternative set of G - 1 equations. In the present study, the first column and then the second column were eliminated. These cases were termed phase 1 and phase 2, respectively.[1]

One of the basic assumptions of the multivariate approach is that off-diagonal elements in Σ are nonzero. (As noted above, if this is not assumed, a univariate approach to estimation and hypothesis testing would be appropriate for each equation.)

This assumption can be put in the form of a null hypothesis to be tested as follows:

$$H: \quad \Sigma = \left\{ \sigma_{gg} \right\}$$

where σ_{gg} is a diagonal matrix. The hypothesis can be tested using a test employed by O'Regan.[2] The test is based upon a statistic calculated from the estimated residual correlation matrix. This statistic is distributed under the stated hypothesis approximately like Chi square with $\frac{G\ (G-1)}{2}$ degrees of freedom, where G is the number of equations.

For the general model presented in this paper, the following methodological approach to multivariate statistical hypothesis testing will be followed.

1/ In the case investigated here, the G - 1 equations in either phase produced identical statistical results (as might be expected).

2/ O'Regan, op. cit., pp. 26 and 27.

Recall the definition of $\hat{\Sigma}$, and now define $\hat{\Sigma}_R$ as the estimated variance-covariance matrix of residuals under restrictions imposed upon the β matrix by the current hypothesis.

Partition the β matrix as follows:

$$\beta = \left[\frac{\beta_1}{\beta_2}\right]$$

where β_1 is K_1 by G and β_2 is K_2 by G (where $K_1 + K_2 = K + 1$).

Consider the set of hypotheses: $\beta_1 = \beta_{1H}$ where β_{1H} is a selected set of constants. The Anderson U statistic is now defined as:

$$U = \frac{|\Sigma|^{N/2}}{|\Sigma_R|^{N/2}}$$

or, more simply, the U statistic can be computed as $\frac{|\Sigma|}{|\Sigma_R|}$, that is, a ratio of determinants.

Recent investigations by Schatzoff substantiate this test statistic selection.[1/] Schatzoff concludes that using the ESL (expected significance level) criterion to judge the six most commonly suggested multivariate test statistics, the two most favorable statistics are the Hotelling T^2 and Wilks' likelihood ratio statistic. This latter statistic is identical to the (unadjusted) Anderson U statistic above. Schatzoff also concludes that these two statistics are asymptotically equivalent.

Now partition the X matrix analogously to that of the β partitioning. The resulting X_1 and X_2 submatrices have the dimensions N by K_1 and N by K_2, respectively.

- -

1/ Schatzoff, op. cit.

The definition of the numerator of the Anderson U statistic is:

$$|\hat{\Sigma}| = |(Y - X\hat{\beta})' (Y - X\hat{\beta})|.$$

The definition of the denominator of the Anderson U statistic is:

$$|\hat{\Sigma}_R| = |(Y - X_1 \beta_{1H} - X_2 \hat{\beta}_2)' (Y - X_1 \beta_{1H} - X_2 \hat{\beta}_2)|.$$

For the specified set of hypotheses, a transformation of the U statistic can be employed. Probability figures for this transformed U can be calculated using the Chi square distribution. The transformation formula is:1/

$$Z = -AM [\log (U)].$$

Log is a natural logarithm, and

$$AM = AN - AK_2 - .5 (G + AK_1 + 1)$$

where

AN = number of observations

AK_2 = number of unrestricted β vectors

AK_1 = number of restricted β vectors

and

G = number of equations.

The statistical routine which calculates the probability figures for the stated hypothesis uses the first two terms of Anderson's suggested formulation. In this general framework, the formula for the β_2 submatrix of coefficients is:

$$\beta_2 = (X_2' X_2)^{-1} [X_2' Y - (X_2' X_1) \beta_{1H}].$$

1/ Anderson, op. cit., Chap. 8.

Application of the Model and Statistical Results Obtained

The multivariate model described above was applied as follows: California Census data on rural farm family income yielded a set of dependent variables.

Each dependent variable consisted of the percent of total income units in a particular income interval.1/

These interval definitions resulted in the estimation of 14 equations for 1950 and 11 equations for 1960. Thus, for the 1950 model, G is equal to 14; and for the 1960 model, G is equal to 11. For both models, the number of observations, N, is equal to 52. Six counties were eliminated because of lack of comparable data over the decade.2/

The explanatory variables might be characterized as of two types: (1) measures of characteristics of farm entrepreneurs (for example, education and age) and (2) measures of characteristics of the spatial unit (for example, index of urbanization, availability of irrigation water, etc.).

In particular, the explanatory variables included the following measures:

1. Real estate--average value of land and buildings per farm. (Values are expressed per $1,000.) This may be a useful proxy for total capital.
2. Education--the recorded average education of farm operators in the county. (Values are expressed in average years of schooling completed.)
3. Farm size--average acreage per farm in the county. (Original values were divided by 1,000.)

1/ For a definition of income intervals, see *supra*, Table 1, p. 6.

2/ The counties eliminated were: Trinity, Sierra, Alpine, Mono, San Francisco, and Inyo.

4. Isard index--a measure of farm-type specialization. This is a modified Isard coefficient of specialization and is discussed in detail below.[1/]

5. Urban index--proportion of the county's population classified as urban.

6. Irrigated farms--proportion of all farms using irrigation water.

7. Road Class 1--proportion of all farms on hard surface roads.[2/]

8. Tenure--proportion of total farms operated by managers.

9. Trading center--a measure of average distance in miles to an urban settlement.

10. Age level--average age of farm operators. (Original values were divided by 10.)

The Census of Agriculture (1950 and 1959) was the basic source of data for the explanatory variables defined above.[3/] The Isard index may be considered in detail at this point. As noted above, the index used here is a modified Isard coefficient of specialization where the Isard coefficient of specialization is defined as:

$$C = \sum_{i}^{N} \left| \frac{M_i}{M} - \frac{S_i}{S} \right| \qquad i = 1, \ldots, N$$

- -

1/ Walter Isard in association with David F. Bramhall et al., Methods of Regional Analysis: An Introduction to Regional Science ("Regional Science Studies Series 4," Vol. II; Cambridge, Massachusetts: The M.I.T. Press, 1960), Chap. 7, pp. 232-308.

2/ Other road classes defined by the Census of Agriculture are: Road Class 2--gravel, shell, or shale; and Road Class 3--dirt or unimproved.

3/ U. S. Bureau of the Census, U. S. Census of Agriculture: 1950. Vol. I, Counties and State Economic Areas, Part 33, California, 1952, 296p.; and idem, U. S. Census of Agriculture: 1959. Vol. I, Counties, Part 48, California, 1961, 287p.

where

C = Isard coefficient

M = total employment in a comparison unit (for example, the entire area analyzed)

M_i = total employment in industry, i, of comparison unit

S = total employment of analyzed unit

S_i = total employment in industry, i, of analyzed unit

i = industry group

and

N = number of industry groups.

For this study, the modified Isard index of farm-type specialization is defined as:

$$C^* = \sum_{i=1}^{N} \left| P_i - \frac{1}{N} \right|$$

where

C^* = modified index

P_i = proportion of county farms which are classified in farm-type i

and

N = number of farm-type classes.

The farm-type classification system used here is that defined by the Census of Agriculture. The following eight major types of farms are defined: commercial, vegetable, fruit and nut, poultry, dairy, livestock, general, and miscellaneous. The objective was to measure the degree of farm-type specialization. If the county had the same number of farms in each of the eight classes, the index value would be zero. The other limit for the index would occur if all

the county's farms were of one type. In this latter case, the index value would be 1.75. These are the lower and upper limits of the index for eight farm types.[1/]

Coefficient estimates were then obtained. (As noted above, singularity led to estimation of the coefficients in two phases, with phase 1 involving the elimination of the first column or dependent variable and phase 2 involving the elimination of the second dependent variable.) Coefficient estimates for 1950 and 1960 appear in Tables 10 and 11, respectively.

The sample residual correlation matrices were calculated for each application of the model, with relatively high values occurring for off-diagonal elements in all cases. As an example, Table 12 lists the residual correlation matrix for phase 1 for 1950.

The hypothesis that the off-diagonal elements equaled zero was tested by the appropriate Chi square test (as described above). Results appear as Table 13.

In each of the cases, the probability that a table Chi square with appropriate degrees of freedom is greater than or equal to the calculated Chi square value is approximately equal to zero. Therefore, each of the hypotheses is rejected, and the multivariate approach is justified for each model.

One method of evaluating the importance of a particular variable in a multivariate framework is to compute the percentage change in the determinant of the variance-covariance matrix of residuals due to that variable.

Computationally, the models presented here result from entering individual explanatory variables in the order which reduces the current $|\hat{\Sigma}|$ the

- -

1/ An index with minimum value equal to zero and maximum value equal to 1 would be constructed by dividing all of these calculated values by 1.75. The value of 1.75 equals $|1 - 1/8| + 7|0 - 1/8| = 7/8 + 7(1/8) = 14/8$.

TABLE 10

Estimated Coefficients, 1950

Constant term and independent variables	Income class interval a/						
	1	2	3	4	5	6	7
	coefficient estimates b/						
Constant term	33.780	9.407	11.050	21.743	35.06	12.240	- 7.039
Urban index	- 5.383	-1.736	- 8.046	- 7.764	- 8.67	- 7.519	5.547
Real estate	- 0.041	-0.015	- 0.020	- 0.053	- 0.03	0.031	- 0.048
Education	- 1.535	-1.475	- 1.153	- 0.635	0.46	0.357	0.093
Tenure (managers)	38.919	2.188	15.163	50.207	114.14	-23.487	71.418
Trading center	- 0.437	0.113	0.587	0.469	0.31	0.072	- 0.239
Road Class 1	- 7.850	-1.290	- 2.280	2.819	3.23	4.443	3.540
Isard index	1.481	-0.531	2.963	3.264	0.72	3.955	- 3.211
Age level	0.896	2.552	2.113	- 1.108	- 4.67	- 1.767	2.070
Irrigated farms	- 2.132	-0.127	1.368	- 0.666	- 2.34	0.413	1.491
Farm size	- 3.552	-1.079	- 3.858	- 0.416	- 0.89	0.702	4.055
	8	9	10	11	12	13	14
Constant term	-17.431	18.829	-17.851	13.12	-13.038	15.184	-15.052
Urban index	8.567	5.925	- 2.236	8.44	4.866	4.433	3.580
Real estate	- 0.017	0.048	- 0.016	0.07	0.017	0.009	0.067
Education	0.839	1.128	- 0.374	0.77	0.123	0.509	0.899
Tenure (managers)	11.906	-70.252	7.671	-134.63	-31.209	-29.444	-22.585
Trading center	- 0.313	0.128	- 0.201	0.12	- 0.133	- 0.465	- 0.006
Road Class 1	- 3.221	0.749	2.874	2.92	- 3.470	- 3.556	1.093
Isard index	- 6.207	2.541	- 0.363	6.06	- 2.385	- 2.267	- 6.026
Age level	3.622	- 6.206	5.202	- 5.53	3.052	- 2.367	2.139
Irrigated farms	1.879	- 1.373	0.101	- 1.07	1.971	1.084	- 0.603
Farm size	1.476	- 0.106	1.674	- 1.37	0.489	1.936	0.936

a/ See <u>supra</u>, Table 1, page 6, for income values included in each class interval.

b/ Number of decimal places varies between income class intervals because of a decision-rule to list data in terms of five significant digits without regard to location of decimal point.

TABLE 11

Estimated Coefficients, 1960

Constant term and independent variables	Income class interval[a]					
	1	2	3	4	5	6
	coefficient estimates[b]					
Constant term	29.091	11.966	14.424	41.227	-1.3140	-32.766
Education	- 0.103	- 1.593	- 1.738	- 1.000	-1.4208	- 0.130
Real estate	- 0.006	- 0.023	- 0.023	0.021	-0.0156	- 0.006
Farm size	- 0.559	1.858	0.206	- 3.209	3.627	3.146
Trading center	- 0.165	0.120	0.304	0.342	0.0684	0.222
Road Class 1	- 2.942	- 5.767	1.695	5.286	8.5510	4.927
Age level	- 2.097	4.918	2.128	- 5.174	3.7886	6.874
Isard index	- 2.421	1.935	- 0.610	3.905	0.4169	- 2.695
Urban index	- 1.850	-12.424	- 0.935	- 2.255	0.3236	2.614
Tenure (managers)	-31.208	5.838	20.835	-48.344	9.6130	13.577
Irrigated farms	- 2.352	0.665	1.823	0.097	-0.8186	4.108

(Continued on next page.)

TABLE 11--continued.

Constant term and independent variables	Income class interval[a]				
	7	8	9	10	11
	coefficient estimates[b]				
Constant term	-16.296	- 2.072	2.929	8.0420	44.768
Education	- 0.530	- 0.189	0.441	1.1089	5.154
Real estate	- 0.008	- 0.022	0.007	-0.0001	0.076
Farm size	- 0.086	3.053	- 0.985	-0.190	- 6.861
Trading center	0.382	- 0.287	0.095	-0.1881	- 0.894
Road Class 1	- 0.845	4.512	- 0.910	-5.5701	- 8.936
Age level	4.639	2.227	0.402	-1.9597	-15.745
Isard index	- 1.782	0.771	- 1.985	-2.7153	5.180
Urban index	6.693	- 3.951	- 0.917	3.1011	9.602
Tenure (managers)	9.980	33.297	-13.681	8.3435	- 8.252
Irrigated farms	0.854	0.334	- 1.737	0.0895	- 3.061

a/ See supra, Table 1, page 6, for income values included in each class interval.

b/ Number of decimal places varies between income class intervals because of a decision-rule to list data in terms of five significant digits without regard to location of decimal point.

TABLE 12

Residual Correlation Matrix, Phase 1, 1950 Model

Income class interval and equation number[a]	Income class interval and equation number						
	2	3	4	5	6	7	8
2	1.0000	0.3890	0.1120	-0.1190	-0.2520	0.0160	-0.0680
3	0.3890	1.0000	0.0930	-0.0450	-0.0690	-0.1650	-0.2960
4	0.1120	0.0930	1.0000	-0.0880	0.0200	-0.0010	-0.2820
5	-0.1190	-0.0450	-0.0880	1.0000	0.0580	-0.3520	0.0150
6	-0.2520	-0.0690	0.0200	0.0580	1.0000	-0.1330	-0.4290
7	0.0160	-0.1650	-0.0010	-0.3520	-0.1330	1.0000	0.1960
8	-0.0680	-0.2960	-0.2820	0.0150	-0.4290	0.1960	1.0000
9	-0.4060	0.0590	0.1150	-0.0290	-0.0800	-0.1490	-0.2450
10	-0.1820	-0.2290	-0.5570	-0.0890	0.0110	0.0710	0.1060
11	-0.3180	-0.0580	0.0080	-0.1210	0.3570	-0.5690	-0.2190
12	-0.3400	-0.1590	-0.2510	-0.0530	-0.1620	0.1330	0.1210
13	-0.0670	-0.2450	-0.3580	0.0390	-0.1710	-0.0550	0.2670
14	0.0490	-0.2520	-0.2410	0.0320	-0.3640	0.0820	0.1710

(Continued on next page.)

TABLE 12--continued.

Income class interval and equation number[a]	Income class interval and equation number					
	9	10	11	12	13	14
2	-0.4060	-0.1820	-0.3180	-0.3400	-0.0670	0.0490
3	0.0590	-0.2290	-0.0580	-0.1590	-0.2450	-0.2520
4	0.1150	-0.5570	0.0080	-0.2510	-0.3580	-0.2410
5	-0.0290	-0.0890	-0.1210	-0.0530	0.0390	0.0320
6	-0.0800	0.0110	0.3570	-0.1620	-0.1710	-0.3640
7	-0.1490	0.0710	-0.5690	0.1330	-0.0550	0.0820
8	-0.2450	0.1060	-0.2190	0.1210	0.2670	0.1710
9	1.0000	-0.2930	0.2110	-0.1100	-0.1980	0.0800
10	-0.2930	1.0000	0.0140	0.4990	0.3130	-0.0290
11	0.2110	0.0140	1.0000	-0.0580	0.1460	-0.3890
12	-0.1100	0.4990	-0.0580	1.0000	0.2720	-0.0400
13	-0.1980	0.3130	0.1460	0.2720	1.0000	-0.0550
14	0.0800	-0.0290	-0.3890	0.0400	-0.0550	1.0000

a/ There are 14 equations corresponding to income class in the model. Because of singularity, one equation must be eliminated. Equation 1 is eliminated in phase 1, which is why the sequence here runs from 2 through 14.

TABLE 13

Results Obtained in Testing the Hypothesis that Off-Diagonal Elements Are Zero in the Variance-Covariance Matrix of Residuals 1950 and 1960

Model[a]	Degrees of freedom	Adjusted calculated Chi square values
<u>1950</u>		
Phase 1	78	171.7704
Phase 2	78	211.1496
<u>1960</u>		
Phase 1	45	151.9073
Phase 2	45	148.4943

a/ In phase 1, column 1 of the Y matrix was excluded; in phase 2, column 2 of the Y matrix was excluded. This was done because of singularity in the original Y matrix.

maximum amount. Thus, in Table 14 each step indicates the amount of residual variation explained by introducing the best variable at that step.

Thus, for the 1950 model the urban index is the first variable to enter, explaining 52 percent of the initial variance to be explained. It is followed by real estate, which explains 43 percent of the residual variance remaining after the urban index is introduced. This, in turn, is followed by education, which explains 39 percent of the variance remaining after the two preceding independent variables are introduced. For the 1960 model the first three variables in order of appearance are education, real estate, and farm size, respectively.

A measure of all currently included variables' effect upon the unexplained variance is the percentage change in the current $|\hat{\Sigma}|$[1/] relative to the determinant of the variance-covariance matrix about the means.[2/] These values are presented in Table 15 for the 1950 and 1960 models, phase 1. This table, in effect, presents the cumulative percent reduction in unexplained variance.

A statistical test of each model was carried out using the methodological approach to hypothesis testing defined earlier. The hypothesis was that all β vectors contain coefficients equal to zero.

Under this hypothesis, the U statistic might be considered as a multivariate analogue of R^2 (in a univariate model, this U statistic is equal to $1 - R^2$). For multiple-equation models, the U statistic does not have exactly the same characteristics as $1 - R^2$ since we are considering the determinant values of matrices of unexplained variance.

- -

1/ "Currently included variables" and "current $|\hat{\Sigma}|$" refer to variables and determinant appearing in a specific step of the stepwise computational procedure.

2/ The "variance-covariance matrix about the means" refers to the variances and covariances of the Y variates before regressions are run.

TABLE 14

Effects of Individual Explanatory Variables Upon a Measure of Generalized Unexplained Variance at Successive Computational Steps, Phase 1, 1950 and 1960

1950 model, phase 1[a]		1960 model, phase 1[a]	
Variables entering at successive steps[b]	Percent change[c]	Variables entering at successive steps[b]	Percent change[c]
Urban index	51.63	Education	44.43
Real estate	43.25	Real estate	41.91
Education	39.03	Farm size	47.23
Tenure (managers)	38.61	Trading center	35.99
Trading center	38.63	Road Class 1	41.19
Road Class 1	51.02	Age level	36.02
Isard index	37.61	Isard index	28.03
Age level	31.90	Urban index	29.96
Irrigated farms	32.77	Tenure (managers)	24.69
Farm size	30.14	Irrigated farms	19.06

a/ Phase 1 involves the elimination of the first dependent variable, Y_1, because of singularity.

b/ One variable enters at each step in the stepwise procedure.

c/ The percent change figure indicates the amount of residual variance explained at each succeeding step. Thus, education (in 1950) explains 39 percent of the variance remaining after the introduction of population level and capital. Variables appear in order of importance in reducing total variance.

TABLE 15

Cumulative Percent Explanation of Initial Generalized Unexplained Variance at Successive Computational Steps, Phase 1, 1950 and 1960

1950 model, phase 1[a]		1960 model, phase 1[a]	
Variables entering at successive steps[b]	Cumulative percent explained[c]	Variables entering at successive steps[b]	Cumulative percent explained[c]
Urban index	51.64	Education	44.43
Real estate	72.55	Real estate	67.72
Education	83.27	Farm size	82.97
Tenure (managers)	89.73	Trading center	89.09
Trading center	93.70	Road Class 1	93.59
Road Class 1	96.91	Age level	95.90
Isard index	98.07	Isard index	97.04
Age level	98.69	Urban index	97.93
Irrigated farms	99.12	Tenure (managers)	98.44
Farm size	99.38	Irrigated farms	98.74

a/ Phase 1 involves the elimination of the first dependent variable, Y_1, because of singularity.

b/ One variable enters at each step in the stepwise procedure.

c/ This measure equals (1 - $|\hat{\Sigma}|$ current/$|\hat{\Sigma}|$ initial) 100 percent where "$|\hat{\Sigma}|$ initial" involves the variance-covariance matrix of the Y variates and "$|\hat{\Sigma}|$ current" involves the estimated residual variance-covariance matrix at a given step in the stepwise procedure.

These tests were carried out for phase 1 for both 1950 and 1960. The unadjusted U statistic under this hypothesis, with 13, 11, and 40 degrees of freedom for the 1950 model, is .00616. The adjusted statistic, Z, with 143 degrees of freedom, is 201.0478. The calculated probability that a table Z value is greater than or equal to the calculated Z value is .00061. For the 1960 model, the unadjusted U statistic, with 10, 11, and 40 degrees of freedom, is .01260. The adjusted statistic, with 110 degrees of freedom, is 179.3393. The corresponding probability is .00002.

Not too surprisingly, then, the hypothesis that all coefficients equal zero was rather resoundingly rejected for all cases investigated.

Hotelling's T^2 statistic can be employed to test the hypothesis that the coefficients of a given independent variable are zero in all equations.[1/] This test was carried out for each of the variables in turn. The T^2 statistics computed for both the 1950 and the 1960 cases appear in Table 16. For each case, Table 16 lists the probability of error if the zero value hypothesis is rejected. For education, Road Class 1, trading center, and real estate, this probability is quite low, indicating the explanatory power of these variables. For age level and irrigation, the probability is relatively high, indicating that these are relatively weak explanatory variables. Tenure, farm size, and the urban index have a low probability in one year, but a relatively high probability in the other, and can be viewed as mixed cases.

Stress was not placed on tests of individual coefficients, given the focus of this study. However, t values for all variables and equations were computed and are presented in Appendix C.

1/ For an explicit definition of Hotelling's T^2 statistic in matrix terms, see Henry Scheffe, The Analysis of Variance (New York: John Wiley & Sons, Inc., 1959), p. 418.

TABLE 16

Calculated Values of Hotelling T^2 Statistic and Error Probability in Testing for Zero Coefficients, 1950 and 1960

Independent variable	1950		1960	
	Hotelling T^2 statistic	Error probability[a/]	Hotelling T^2 statistic	Error probability[a/]
Education	45.39*	.0089	71.57*	.00002
Road Class 1	74.87*	.0002	32.30*	.0130
Tenure (managers)	38.12*	.0260	15.41	.2780
Age level	27.83	.1160	16.63	.2280
Trading center	50.32*	.0044	38.42*	.0042
Irrigated farms	24.84	.1760	12.25	.4520
Farm size	22.43	.2440	40.50*	.0029
Real estate	51.10*	.0039	46.25*	.0010
Urban index	57.77*	.0015	20.91	.1070
Isard index	30.03	.0850	26.22*	.0404

a/ Probability of error if the null hypothesis is rejected. The null hypothesis states that the variable has zero coefficients in all equations.

* Statistically significant at the 5 percent level.

Interpretations of Results Obtained

The estimated coefficients obtained are plotted against the midpoints of their corresponding income class intervals in Figures 6 through 15. For a particular variable of a given model, the relation between the estimated coefficients and income level is shown for both 1950 and 1960. Thus, each graph summarizes the results for that variable in 25 equations (14 for 1950 and 11 for 1960).

Coefficients are indicated as points. Straight-line approximations are drawn between successive points; 1960 coefficients are plotted against appropriately deflated values. Since 1950 has 14 classes and 1960 has 11 classes, there will be some lack of correspondence in the comparisons. Further, the plot results are not adjusted for different class sizes. This lack of adjustment for class size is defensible in terms of exhibiting coefficient values actually obtained. However, for a comparison between classes of the impact of a unit change in an independent variable, it would be sensible to take differing class sizes into account. To put this in simple terms, say a particular coefficient is .1 for income Class A and .3 for income Class B; then, an increase of 10 units in the independent variable would lead to a predicted increase of one income unit in A and three income units in B. However, if it were further specified that the Class B interval is three times as large as that of A, the increase in income units per thousand dollars spanned by the interval will be the same for B as for A. This effect should be kept in mind in interpreting results.1/

In interpreting results, the following general considerations apply. A variable with a positive coefficient increases the frequency of occurrence in

---

1/ The open-end income interval, in an adjusted graphing, would have to be treated as a separate category.

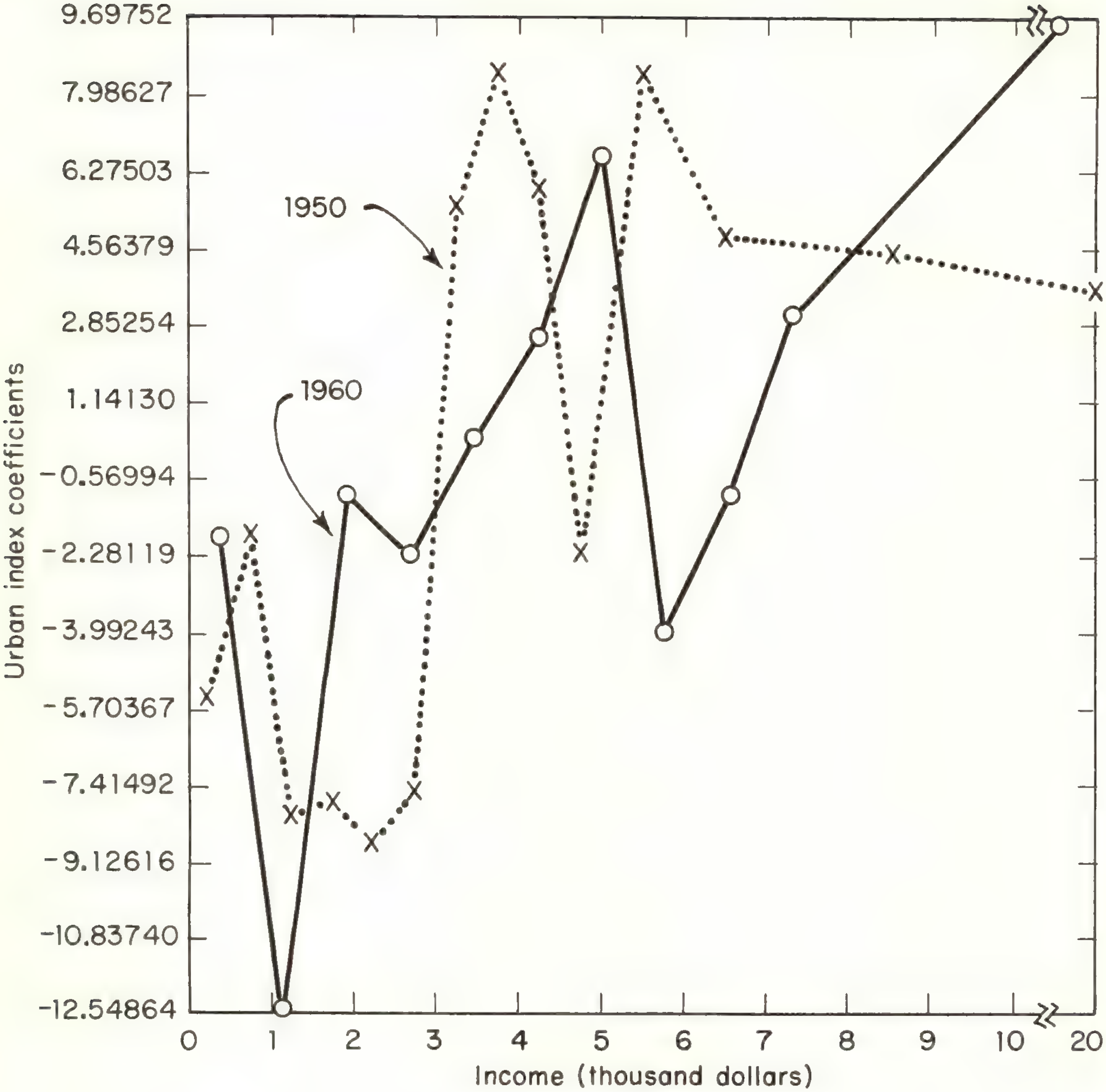


FIGURE 6. Plot of Estimated Coefficients for Urban Index on Midpoints of Income Classes, 1950 and 1960

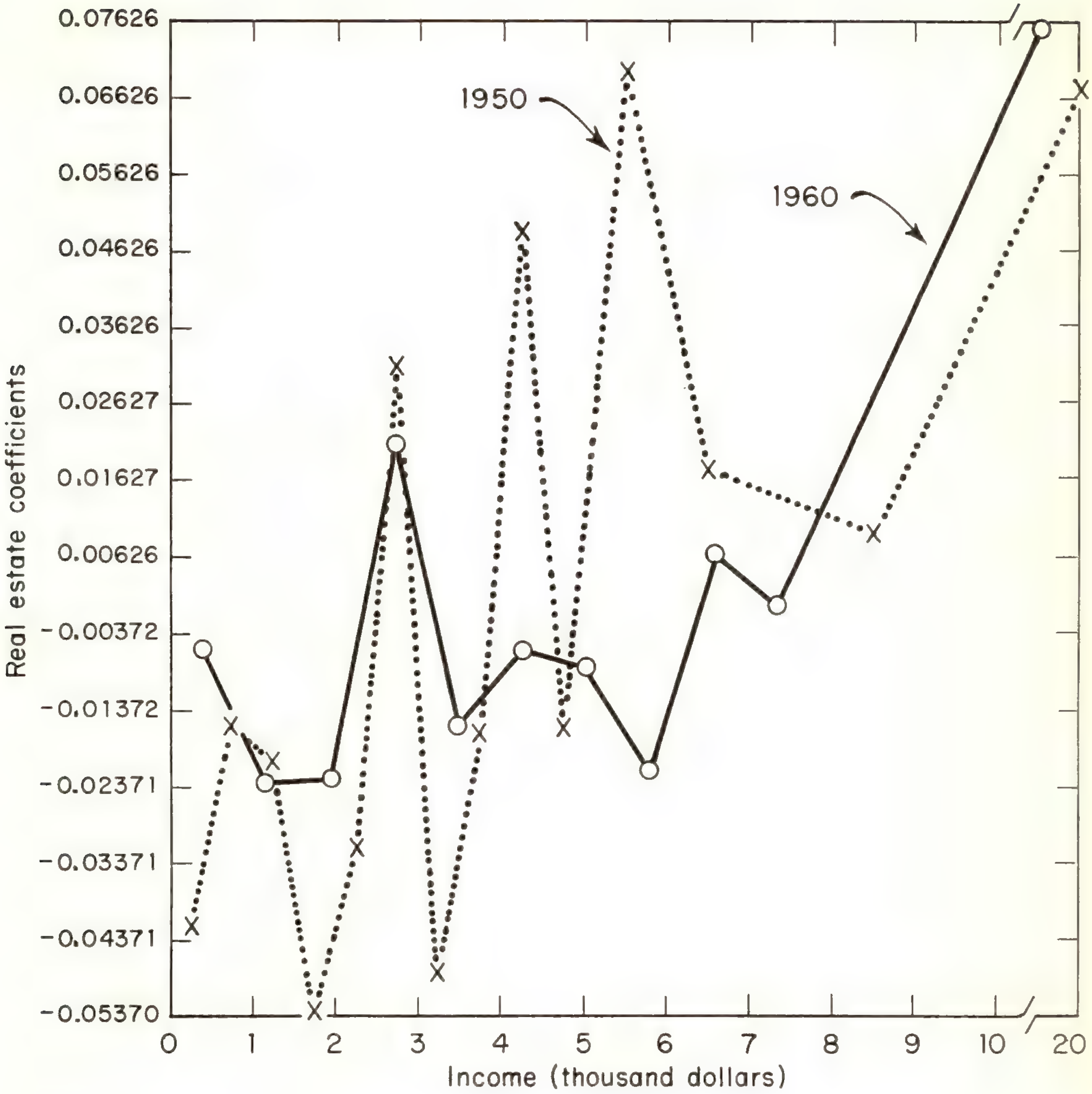


FIGURE 7. Plot of Estimated Coefficients for Farm Real Estate on Midpoints of Income Classes, 1950 and 1960

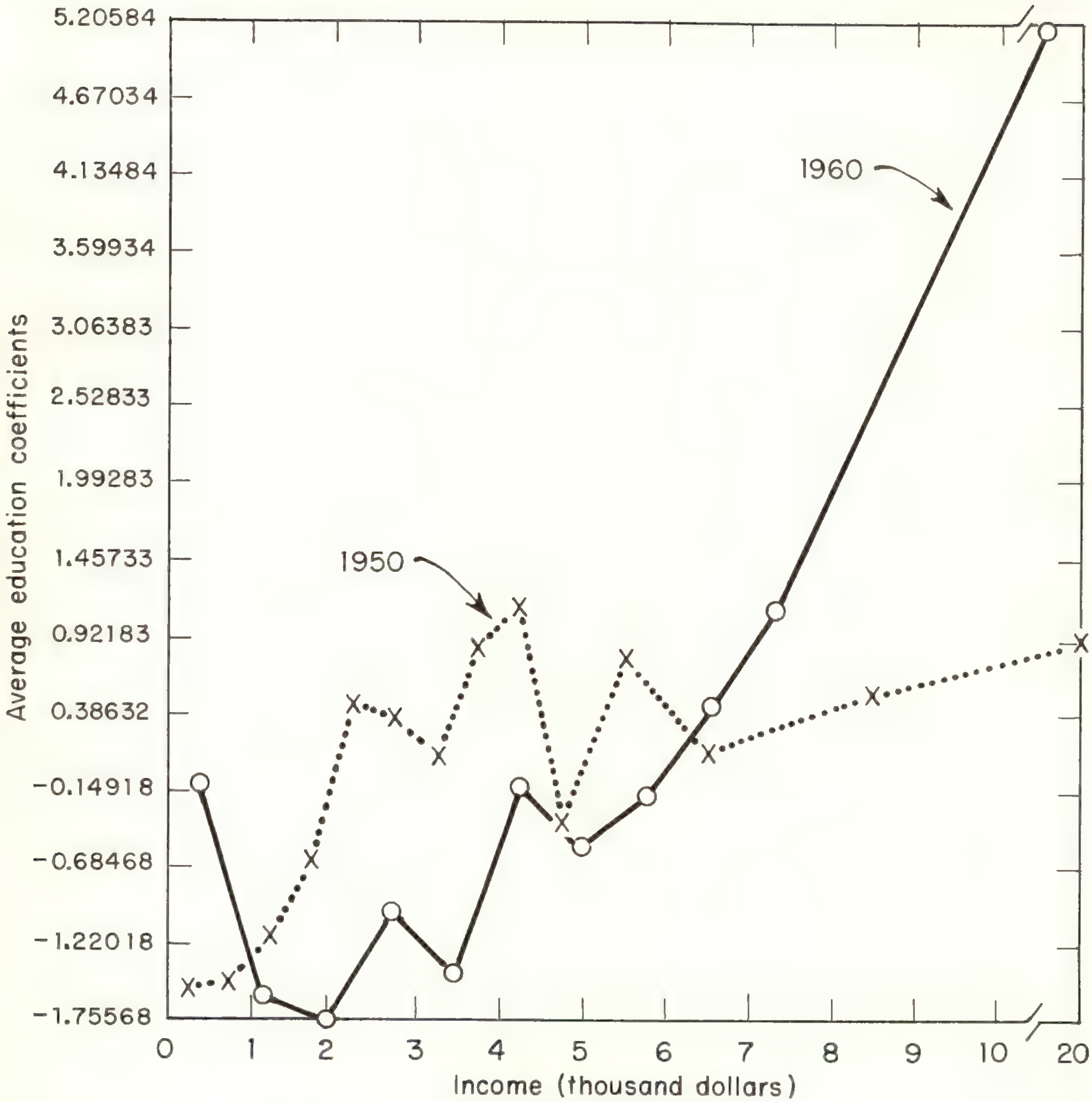


FIGURE 8. Plot of Estimated Coefficients for Average Education on Midpoints of Income Classes, 1950 and 1960

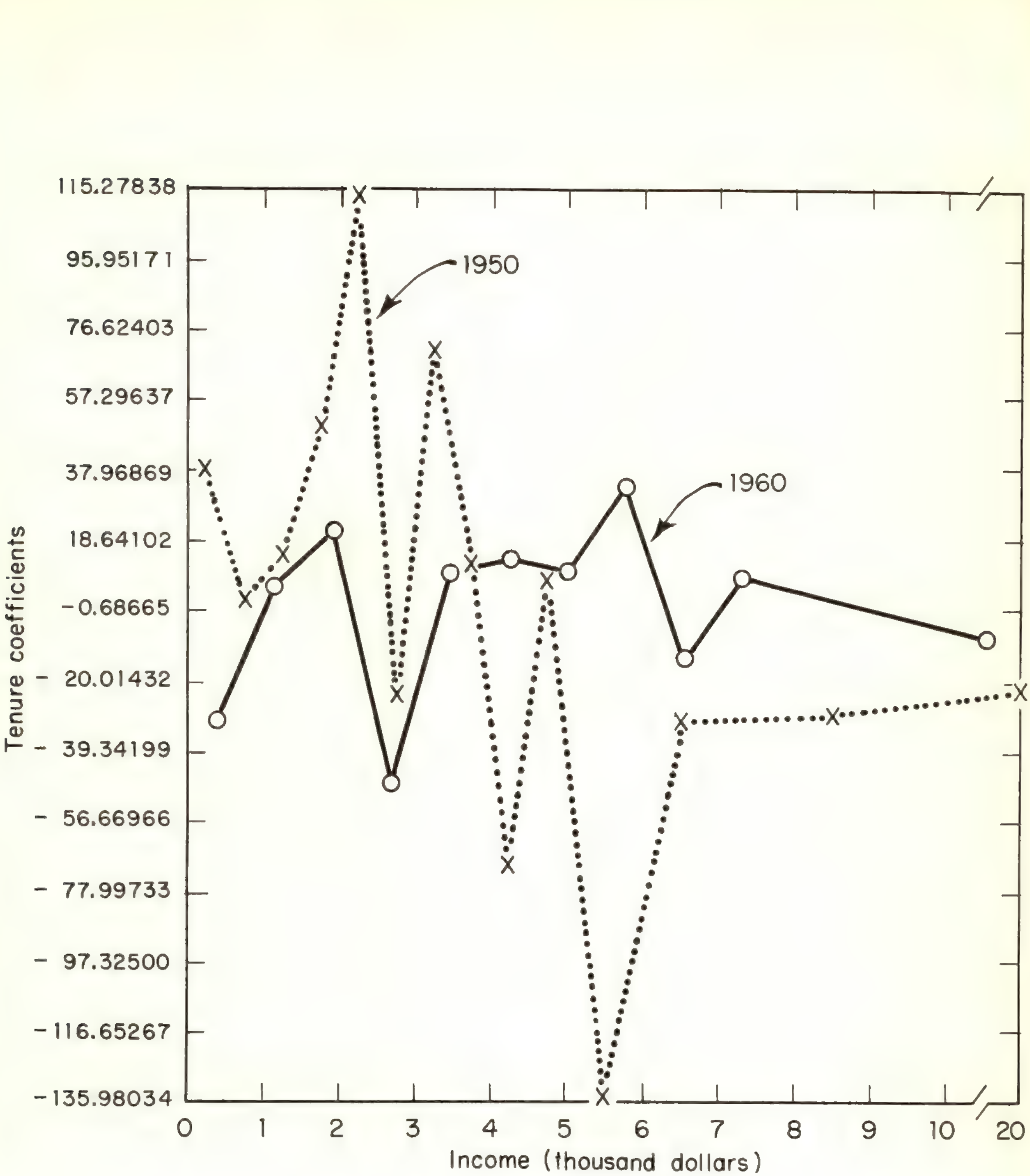


FIGURE 9. Plot of Estimated Coefficients for Tenure (Proportion of Hired Managers) on Midpoints of Income Classes 1950 and 1960

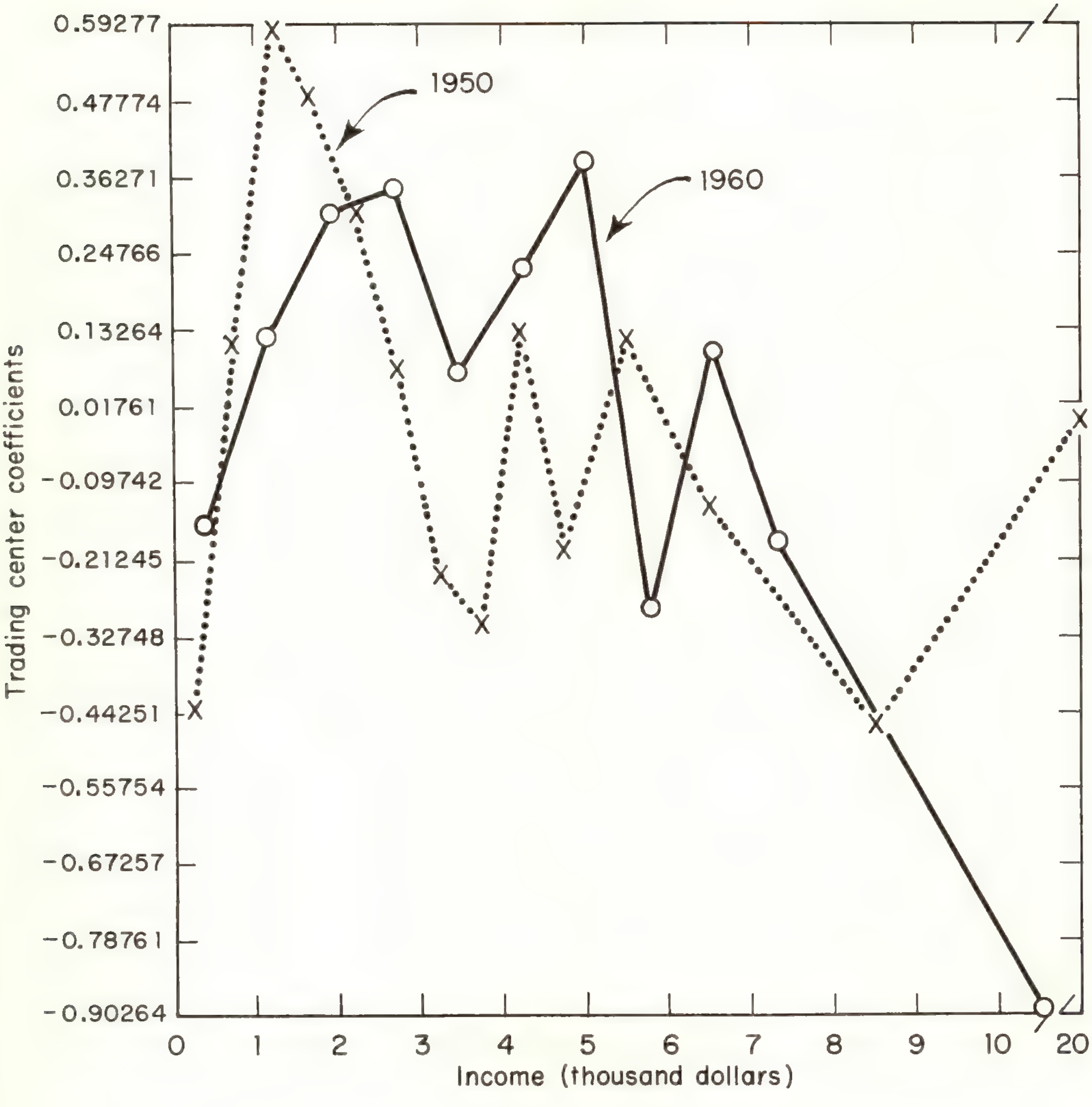


FIGURE 10. Plot of Estimated Coefficients for Trading Center (Distance to Urban Settlement) on Midpoints of Income Classes, 1950 and 1960

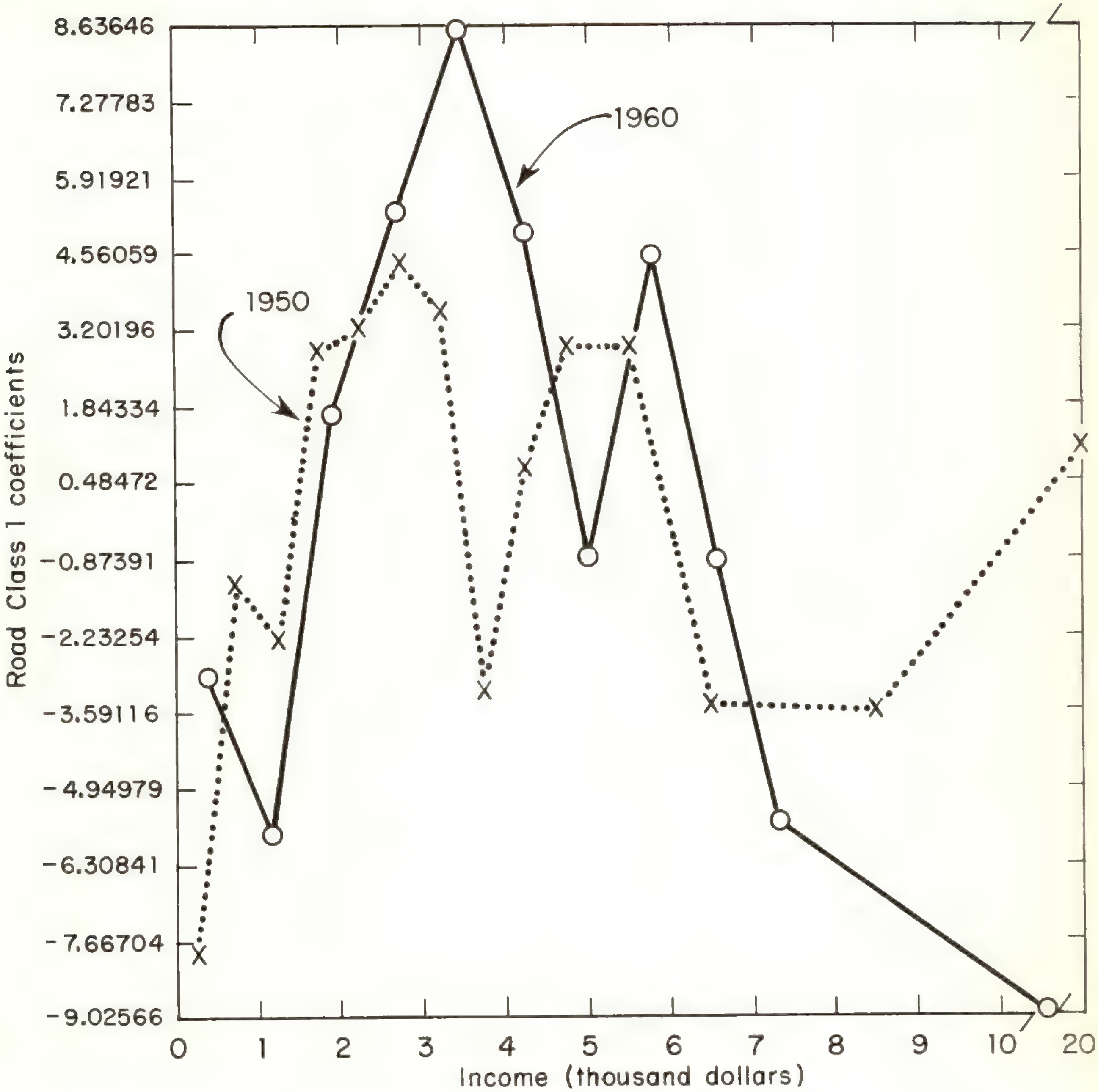


FIGURE 11. Plot of Estimated Coefficients for the Proportion of Road Class 1 on Midpoints of Income Classes 1950 and 1960

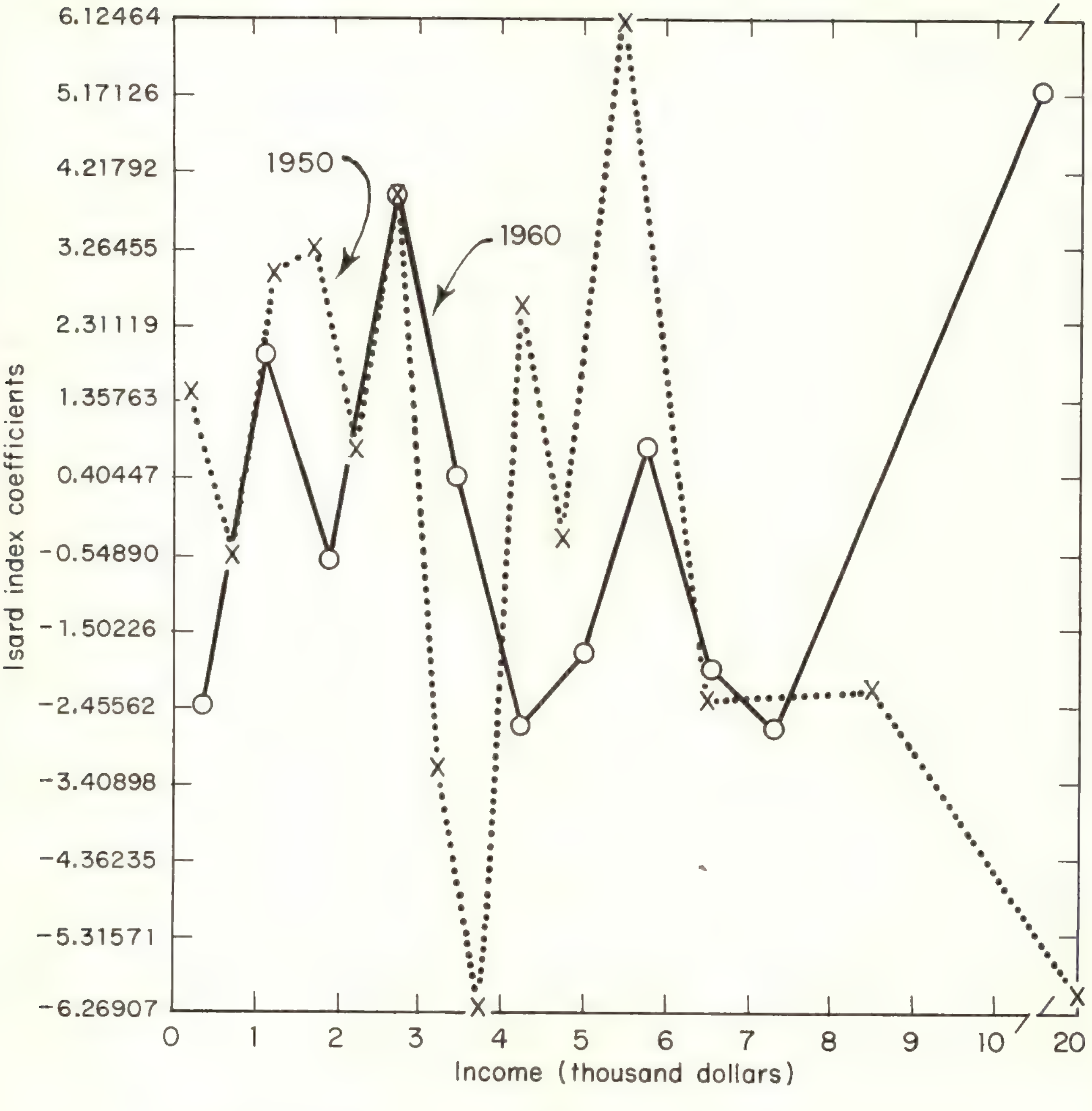


FIGURE 12. Plot of Estimated Coefficients for the Isard Index on Midpoints of Income Classes, 1950 and 1960

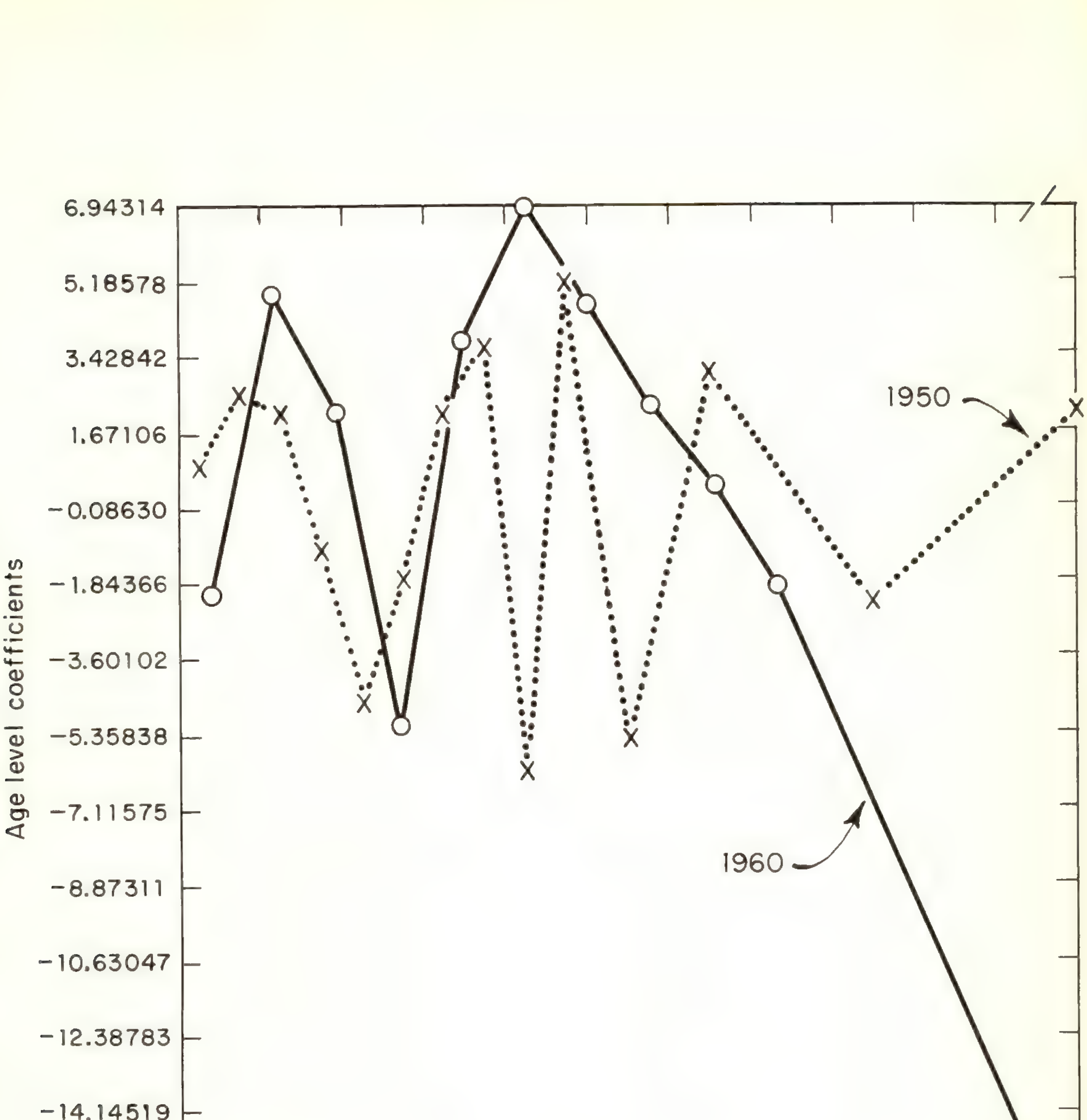


FIGURE 13. Plot of Estimated Coefficients for Average Age on Midpoints of Income Classes, 1950 and 1960

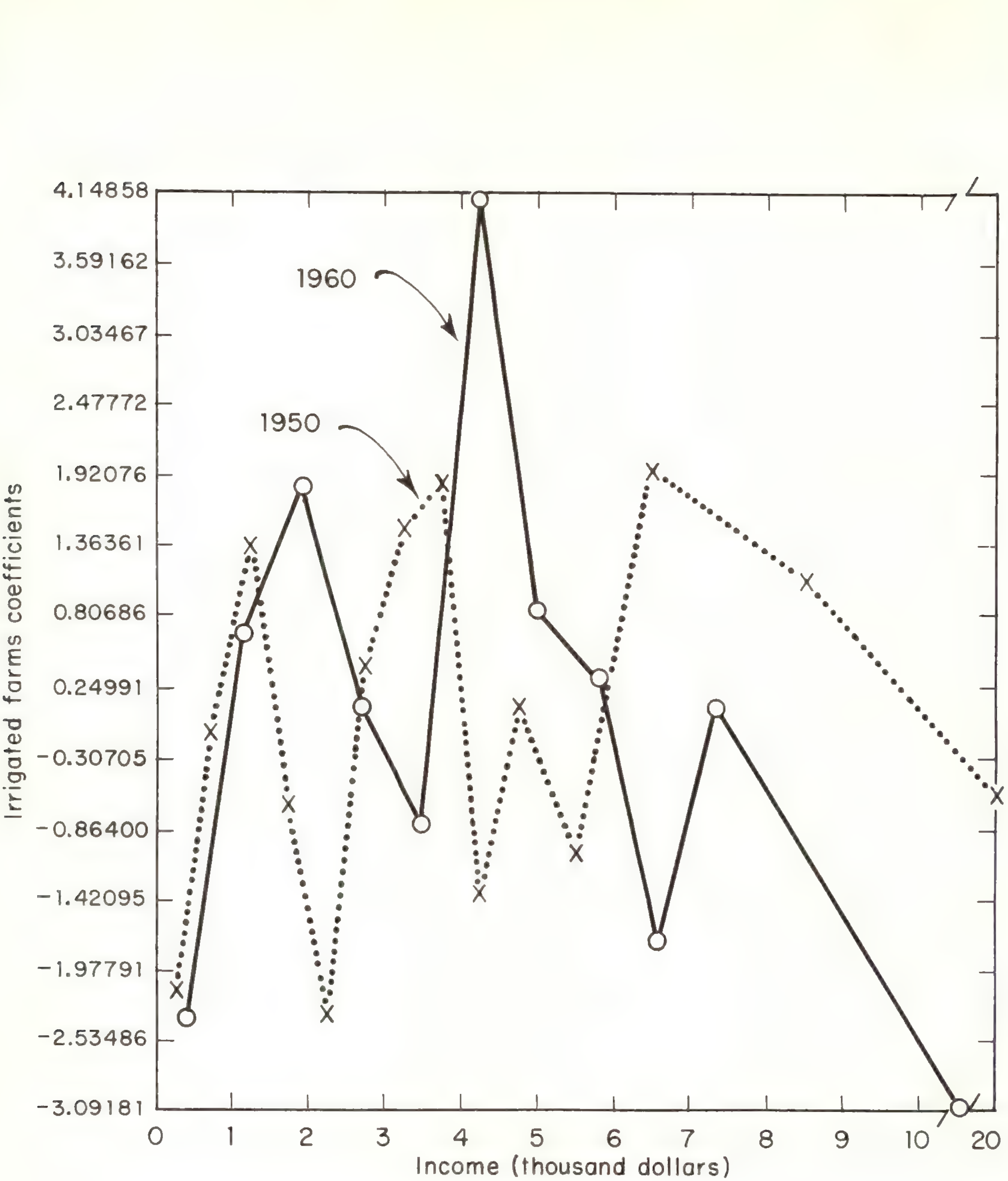


FIGURE 14. Plot of Estimated Coefficients for the Proportion of Farms Irrigated on Midpoints of Income Classes 1950 and 1960

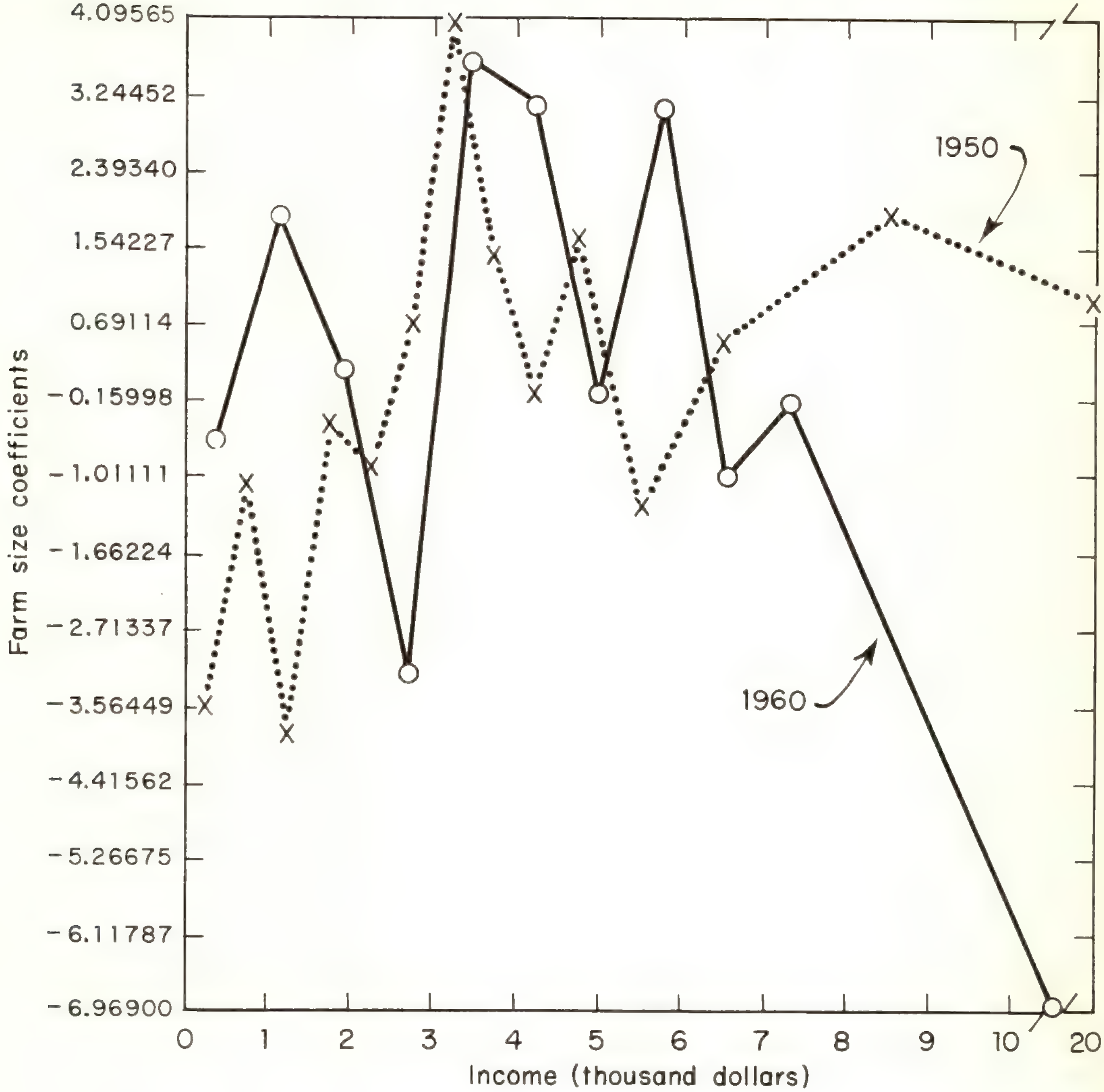


FIGURE 15. Plot of Estimated Coefficients for Farm Size on Midpoints of Income Classes, 1950 and 1960

the given income class, and a negative coefficient decreases the frequency of occurrence. If a variable exhibits an increasing coefficient as income increases, moving from negative to positive values, then this indicates that increasing the value of the explanatory variable will decrease the number of low-income recipients and increase the number of high-income recipients. The reverse relation occurs, of course, if the coefficient decreases as income increases. Other patterns are of interest. Thus, a coefficient may be negative at the extremes of the income distribution and positive in the middle income classes. An increase in the associated independent variable yields greater income equality, though it may have little impact on the average level of income. (This information can be viewed as one of the advantages of working with a distribution rather than with averages.)

Implicit in this discussion is a general feature of the coefficient results which is of considerable interest. In Tables 10 and 11, the sum of the constant term over income classes is extremely close to 100, corresponding to a total frequency of 100 percent. The sum of the coefficients of each dependent variable over income classes is essentially zero. No doubt this reflects the very large amount of explained variance as shown in Table 15.[1/]

A consequence of this feature is that a change in an independent variable yields a new income distribution with income class frequencies summing to the proper 100 percent. A further consequence is that each independent variable has at least one positive and one negative coefficient over the set of all income classes. Finally, coefficients for a subset of income intervals can be

- -

1/ A similar result occurred in some preliminary work using number of income recipients, rather than frequencies, as dependent variables. In this case, the total number of income recipients appeared as one of the independent variables; results involved essentially a zero sum of coefficients for the other independents, a sum of 1.00 for the coefficients of total recipients, and a sum of zero for the constant term.

aggregated so the impact of a change in an independent variable can be viewed in broad terms. For example, in the 1950 case, an increase of one year in the education level has the following impact, based on Table 10 data. There is a drop of 4.8 percent in income recipients receiving \$2,000 or less; an increase of 2.5 percent in the \$2,000-\$5,000 class; and an increase of 2.3 percent in the over-\$5,000 class.

Turning to results for individual variables, the coefficient patterns for the urban index, real estate, education, and trading center seem to emerge rather clearly and fit expectations rather well. Coefficients for the latter variable decline as income increases. This is reasonable, for an increase in distance to trading center leads to an increase in low-income recipients, a decrease in high-income recipients, and, hence, a decline in income overall. For the other variables listed above, there is a general upward movement in coefficient value as income increases, indicating a positive shift in income distributions as these variables increase. The positive impact of urbanization adds some further support to the location hypothesis of T. W. Schultz.[1/]

The education coefficients show a clearly increasing pattern in both 1950 and 1960. However, some shift in pattern appears involved, with the 1960 education impact seeming somewhat more pronounced. Such results appear meaningful in considering education as an income shifter. One of the difficulties with prior measures of the impact of education is that there is no guarantee that the underlying population is homogeneous. Thus, it may be found that income increases with education; however, if there are differences in innate ability involved, both increases may reflect this latter factor. In the present study, however, this problem seems much less likely to occur. It seems plausible to

- -

1/ Schultz, op. cit.

argue that there is not much difference in average level of innate ability between counties. Hence, the impact of the education variable can be attributed plausibly to the education process with much less concern that differences in underlying ability are involved. Of course, average education level involves an underlying distribution; and, hence, some improvement for both analysis and policy purposes can be expected by disaggregation. Thus, in future work the education variable might be replaced by a set of education activities, including primary, secondary, and college education; level of extension work; and measures of "quality" of education.

The real estate coefficients exhibit an upward movement with income, though the 1950 estimates show rather pronounced variability. Results for farm size show a different pattern. There is little trend in value of coefficient as income increases. This is the case for both 1950 and 1960, with one exception occurring in the 1960 results; here, there is a pronounced drop in coefficient value for the highest income class. It turns out there is little correlation between real estate, measured in dollars per farm, and farm size, measured in acres per farm. (This correlation was about .15 in both 1950 and 1960.) It is likely that very large acreages are associated with grazing operations so that the farm size variable may be an index of grazing; such an argument might be applicable in explaining results for this variable.

Results for Road Class 1 indicate that good roads lead to greater equality of income. Thus, for both 1950 and 1960, increasing the level of this variable reduces income recipients in both the lowest three and the highest three income classes. In 1950 the reductions are 11 and 6 percent, respectively; while in 1960, the respective reductions are 7 and 15 percent. Before road building is recommended seriously as an egalitarian device, however, it is worth noting that this variable shows a fair amount of negative correlation with distance to

trading center--which seems reasonable. The correlation between these variables was -.5 in 1950 and -.6 in 1960. The building of good roads will tend to reduce distance to trading center, both in terms of over-the-road mileage and by possible generation of new trading centers. Hence, it might be best to consider these variables jointly in interpreting results.

Results for the remaining four variables are not particularly enlightening. The variable labeled "tenure" is a measure of the proportion of hired managers. For 1950 the coefficients exhibit some decline as income increases, but little pattern appears for 1960. The Isard index, which measures specialization, shows a great deal of variability and little pattern in either 1950 or 1960. The age variable shows little pattern for 1950 but in 1960 exhibits a marked declining trend for upper income groups, indicating some income decline with increasing age. Future work might allow for a possible nonlinear relation between age and income; this could be done by using several age variables or by introducing the squared value of age minus a constant. Finally, the irrigated farm variable, measured as a fraction of all farms, shows little in the way of pattern for both 1950 and 1960. Future work might experiment with alternative measures, such as irrigated acreage as a fraction of all acreage.

Overall, there is a good deal of similarity in the pattern of behavior of a given coefficient for 1950 and 1960. Most of the variables exhibit roughly the same graphic relationships in the two periods. However, though patterns seem intuitively similar, evidence to be presented below indicates some changes in structure may have taken place between 1950 and 1960.

The evidence emerges as a by-product of a specific application. The results developed here can be applied in forecasting income distributions if (1) forecasts of explanatory variables are made and (2) adjustments for price level changes and underreporting are carried out.

By way of example, forecasts were developed for the distribution of income for rural farm families in Fresno County. Fresno County was selected because it is the largest county in terms of aggregate rural farm income. The 1950 equations were used to develop Fresno County forecasts for 1950 and 1960. The 1950 "forecast" can be viewed as indicative of the goodness of fit of results for a specific county in 1950.

Forecasts presented here are in terms of the 1950 income class intervals. Table 17 compares the 1950 forecast to 1950 actual for each income class.

On an intuitive level, forecasts seem fairly close to actual levels. Thus, 8 of the 14 forecasts are within 10 percent of actual, 3 are between 5 and 10 percent, and 3 are above 10 percent. Rather large discrepancies occur for the extreme income classes (the lowest and highest).

Comparison of 1960 forecasts to 1960 actual values is complicated by underreporting of income in 1950 relative to 1960 and by increases in the general level of prices. These problems were handled by adjusting income intervals and then redistributing number (or frequency) of income recipients to the income classes defined for 1960. In particular, available evidence indicated a price inflation factor of 1.223 and a general income underreporting factor of 1.1, with greater adjustment occurring for lower income classes.[1/]

Results obtained appear in Table 18, which compares 1960 forecasts to 1960 actual values for Fresno County. It turns out that forecasts are well above actual values for lower income classes and well below actual values for

---

1/ The price inflation factor was obtained from the consumer price index for 1959 relative to 1949; the income underreporting factors from Irving Hoch, Income and Taxes (Chicago, Illinois: Chicago Area Transportation Study, 1958), p. 31. The approach followed involved the assumption that adjustments were uniform within an income class; as a consequence, class intervals were modified. Thus, the 1950 income class interval, \$1,000-\$2,000, was transformed to an interval of \$1,223-\$2,446 on the basis of the price inflation factor of 1.223.

TABLE 17

Forecasts of Income Distribution Using the 1950 Coefficients
Fresno County, 1950 and 1960

Income class: Number	Income class: Interval: Range from lower value to less than upper value	Frequency of rural farm families falling in income class[a]: 1950 actual	1950 forecast	1960 forecast	1950 forecast divided by 1950 actual
	thousand dollars				
1	0- 0.5	.0717	.0820	.0742	1.144
2	0.5- 1.0	.0632	.0633	.0571	1.002
3	1.0- 1.5	.0884	.0868	.0816	.982
4	1.5- 2.0	.1034	.1053	.0959	1.018
5	2.0- 2.5	.1299	.1217	.1159	.937
6	2.5- 3.0	.1046	.0942	.0993	.901
7	3.0- 3.5	.0942	.0924	.0885	.981
8	3.5- 4.0	.0666	.0635	.0617	.953
9	4.0- 4.5	.0579	.0569	.0674	.983
10	4.5- 5.0	.0414	.0400	.0370	.966
11	5.0- 6.0	.0539	.0717	.0869	1.330
12	6.0- 7.0	.0330	.0325	.0365	.985
13	7.0-10.0	.0539	.0488	.0506	.905
14	10.0+	.0378	.0419	.0486	1.108
	Total	1.0000	1.0010	1.0012	1.001

a/ Frequency, in terms of proportions.

TABLE 18

Comparison of 1960 Forecasts (Adjusted for Underreporting) to 1960 Actual Income Distribution, Fresno County

Income class: Number	Income class: Interval: Range from lower value to less than upper value	Frequency of rural farm families falling in income class[a/]: Actual	Frequency of rural farm families falling in income class[a/]: Forecast	Forecast divided by actual
	thousand dollars			
1	0- 1	.0628	.0828	1.318
2	1- 2	.0757	.0852	1.125
3	2- 3	.1120	.1540	1.375
4	3- 4	.1456	.1921	1.319
5	4- 5	.1182	.1225	1.036
6	5- 6	.1087	.0996	.916
7	6- 7	.0794	.0598	.753
8	7- 8	.0574	.0652	1.136
9	8- 9	.0412	.0298	.723
10	9-10	.0342	.0180	.526
11	10+	.1648	.0921	.559
1- 5	0- 5	.5143	.6366	1.238
6-11	5-10+	.4857	.3645	.750
	Total	1.0000	1.0011	1.001

a/ Frequency, in terms of proportions.

upper income classes. The magnitude of forecast error is about 25 percent, given an aggregation into high- and low-income classes. These results suggest a change in structure occurred between 1950 and 1960.

Conclusion

The first major section of this report examined patterns of rural farm income on a county basis. At the methodological level, one of the techniques employed was the classification and mapping of income levels by means of a computer program. It seems likely that there will be considerable growth in the use of such computer mapping techniques in regional analysis. In terms of substantive results, there is some support for the following conclusions. For the state of California, no marked difference in income level occurs between farm and nonfarm sectors at the county level. Urbanization is associated with increased income levels for both nonfarm and farm populations. Some reduction of income differentials appears to be occurring at the county level; that is, counties with lowest average income tend to have the highest rate of growth in average income. Nevertheless, marked income differentials persist over time. For the farm sector, seven counties had high average incomes and five counties had low average incomes in both 1950 and 1960, using a high, medium, and low classification. In the nonfarm sector, 9 counties were in the high classification and 11 were in the low classification in both periods. Thus, the "income problem," in terms of pronounced differences in average level between groups, seems more of a regional than a sectorial problem in California. Finally, there appear to be regional clusters of relatively equal income growth. For both farm and nonfarm sectors, southern California, the southern coast, and a tier of counties north of Sacramento County increased markedly in average income rank. At the other extreme, San Joaquin Valley and northern California counties declined substantially in average income rank.

The work of the second section can be viewed as an attempt to uncover some of the underlying causative factors explaining the income differentials discussed here. But beyond that kind of work, deeper probing and alternative explanatory frameworks could be considered. This could involve such matters as the persistence of income differences over long periods, their relation to migration and population growth, and the effectiveness of policies aimed at reducing income differentials.

On the methodological level, the second major section of this report applied multivariate analysis to the study of income distributions. For each income class, the frequency in that class was related to a common set of explanatory variables. Good substantive results were obtained in terms of explained variance and coefficient estimates for a number of the explanatory variables. In particular, results for variables--education, real estate, urbanization, and trading center--seemed reasonable and in line with expectations. A noteworthy aspect of the estimates is that, for each independent variable, the coefficient sum over income classes totals zero, while the sum of constants totals 100. As a consequence, any change in the levels of explanatory variables leaves total frequency unaffected. This is a useful feature both in terms of viewing the impact of a unit change in one independent variable and in terms of obtaining new distributions by changing levels of all independent variables, given estimates of those variables. Estimating levels of independent variables is a feature of two important applications: (1) obtaining forecasts for specific counties (as exemplified here in the case of Fresno County) and (2) estimating income distributions for areas smaller than counties. The latter application may have commercial possibilities. Thus, consider a firm marketing a consumer product with a high income elasticity of demand; information on small region income distributions should be of help in its sales and marketing strategy.

Some avenues of future work may be noted. First, a number of new independent variables might be introduced. In particular, some suggested variations for education, age, and irrigation were noted above. Second, attempts can be made to investigate the underlying structural relationships which are roughly indicated by the figures plotting estimated coefficients against income class midpoints. An attempt might be made to develop smooth curves for the relationships and, along with this, to test the hypothesis that a change in structure has occurred between time periods. The results obtained for the Fresno forecast indicate that such a change has in fact occurred. It may be that part of the discrepancy involves a statistical problem stemming from the impact of transitory (versus permanent) components of income. But, in any event, it seems plausible that some underlying changes in structural relationships have in fact occurred. This is suggested, in particular, by the shift in pattern noted for education between 1950 and 1960. Finally, there are intimations, at least, that the analysis developed here can be developed further for use as a policy tool. The explanatory variables would then be viewed as policy instruments. In particular, they might be used to achieve not only higher levels of average income but a "better" distribution of income, presumably reflecting the views of the electorate.

APPENDIX A

Problems and Solutions in the Use of Census Income Data

A number of problems arise in the use of Census of Population income data in economic analysis. Problems handled in this study include (1) the distribution of nonresponse units to make 1950 data comparable to 1960 and (2) the development of average income estimates accounting for skewness in the distributions. Other problems noted in this appendix are (1) the problem of differences in U. S. Office of Business Economics and Census measures of income and (2) the problem of comparability of family income versus family and unrelated income. The discussion of these problems can be of relevance in applications of results presented in the body of this report.

Nonresponse

An important difference between 1950 and 1960 Census income data is in the handling of nonrespondents. Of the sample interviewed, the nonrespondents for income data are those units for which some characteristics were recorded but income data were not obtained. Nonresponse occurred either because the interviewee did not have adequate information or because he refused to divulge it.[1/]

The 1950 Census listed the number of nonresponse units as a separate category. In the 1960 Census, units in this category "were assigned the reported income of persons with similar demographic and economic characteristics."[2/]

1/ Personal correspondence with Henry S. Shryock, Jr., Acting Chief, Population Division, U. S. Bureau of the Census, March 12, 1965.

2/ Ibid.

The 1960 Census used the following characteristics for distributing the nonrespondents: age, sex, family status, color, weeks worked, and major occupation group.1/

To increase comparability between the 1950 and 1960 income distribution data, a procedure was developed for distributing the 1950 income nonrespondents in a manner similar to that used by the Census in 1960.

Unpublished data from the Bureau of the Census were employed in the distribution procedure.2/ A portion of this data is contained in Appendix Table A-1. This table shows the 1960 income distribution of male persons (14 years old and over) before and after distribution of nonrespondents for the state of California. The income distribution of male nonrespondents, as allocated by the 1960 Census procedure, is shown. In addition, the ratio of the number "after" to the number "before" for each income class is shown.

The procedure for distributing nonrespondents consisted of the following steps:

1. A matrix of "weights" for the adjustment of each income matrix was computed. (Counties correspond to rows and income classes to columns.)
2. These weights were multiplied by the observed number of units in each income class interval for each county.
3. The above products were summed over income classes of the respective county. The resulting sum was then compared to the actual sum of observation units for that county.

1/ For a brief description of the 1960 distributing of income nonrespondents, see U. S. Bureau of the Census, U. S. Census of Population: 1960 . . ., 1963, pp. XXVIII and XXIX.

2/ Personal correspondence with Henry S. Shryock, Jr.

APPENDIX TABLE A-1

Male Income Distribution Before and After Distribution of Nonrespondents, 1960

Income class: Interval: Range from lower value to less than upper value (thousand dollars)	Nonrespondents	Before distribution of nonrespondents	After distribution of nonrespondents	Ratio of after to before
0- 1	49,573	504,330	553,903	1.098
1- 2	42,300	528,895	571,195	1.080
2- 3	31,404	385,246	416,650	1.082
3- 4	32,740	424,875	457,615	1.077
4- 5	37,362	517,745	555,107	1.072
5- 6	38,538	618,267	656,805	1.062
6- 7	31,110	544,885	575,995	1.057
7-10	41,175	763,505	804,680	1.054
10+	26,588	454,396	480,984	1.058
Total	330,790	4,742,144	5,072,934	1.070

Source: Unpublished data obtained by personal correspondence with Henry S. Shryock, Jr., Acting Chief, Population Division, U. S. Bureau of the Census, March 12, 1965.

4. When the two sums were not equal, each of the adjusted number of units in each income interval for each county were multiplied by the ratio of the actual sum to the computed sum.

The result for each county of the above four adjustment steps is to distribute the 1950 income nonrespondents over that county's 1950 income distribution in a manner similar to that used by the 1960 Census. Consequently, the sum of the resulting distribution is equal to that of the county's income respondents plus nonrespondents.

To restate the procedure in algebraic terms, the number of observations in a given income class interval for a given county is multiplied by a scale factor:

$$W_{spij}$$

where

s = sector

p = income unit (either family or family and unrelated)

i = county

and

j = income class interval.

Then, W_{spij} is defined as follows:

$$W_{spij} = \frac{(R_j)\ (R_{t,\ spi})}{(R_t)}$$

where

R_j = the state of California ratio of number of observations after distributing nonrespondents to the number of observations before distributing nonrespondents by income class, j. (Numerical values of R_j appear as the last column of Appendix Table A-1.)

R_t = the ratio of after to before of total California observations (and equals 1.070 as shown in Appendix Table A-1)

and

$R_{t,\ spi}$ = the ratio of total county, i, observations including nonrespondents to the total excluding nonrespondents for specific sector, s, and income unit, p.

Estimating Average Income for Skewed Income Distributions

Given the adjusted observations on county income distributions, average income estimates by county and sector were then developed.

Average income is often computed by cumulative summing of income units times midpoints of class intervals and dividing by total units. However, since income distributions are usually positively skewed, this approach was not used here. Instead, average income was estimated as a function of the midpoint estimate and of a measure of skewness computed under the midpoint assumption.1/ In the application of this function, the midpoint estimate and skewness measure were developed for each distribution of this study. The coefficients of the

- -

1/ Alternative approaches were considered but posed difficulties. The skewed distribution is often transformed to approach a symmetric distribution. The usual transformation procedure for income distributions employs the log transformation. This is correct when X is the positive population variable and Log (X) = Z is in fact normally distributed. This approach is convenient and easily adopted. However, it is not appropriate in transforming the various forms of income distributions which appear on a sectorial breakdown.

An alternative transformation technique is to employ the Sech square distribution function. This approach can be applied to a wide range of distribution forms. Although application is of a simple nature, a significant amount of decision-making time is required in a single application. See Peter R. Fisk, "The Graduation of Income Distributions," Econometrica, Vol. 29, No. 2 (April, 1961), pp. 171-185.

function were given, having been estimated previously using United States data. Three forms of the function were employed, with a separate set of coefficients given for farm operator families, nonfarm families, and all families. (The results for farm operator families were assumed to be applicable to the rural farm family units here.[1/])

In estimating the functions to be employed, data used were United States family data collected and published by the Office of Business Economics (OBE).[2/] A nine-year time series was available on three sectors. (This included the years 1950, 1951, and 1953 through 1959.) As indicated above, the data pertained to farm operator families, nonfarm families, and all families as defined by the OBE and assumed applicable here. Value deflation was carried out, using the consumer price index. Differences do exist between the OBE definitions of sectors and observation units and the Census definitions. But this was not considered to be of critical importance for the problem at hand. The most obvious assumption involved is that aggregated United States sectorial data are roughly appropriate for correction of moments computed from California sectorial data.

It was hypothesized that the following relations held:

$$\frac{Y_{mi}}{Y_{ai}} = A_o + A_1 (S_{ai})$$

$$S_{ai} = B_o + B_1 (S_{mi})$$

- -

1/ For the open-end upper income classes, it was assumed that the class average for the "$10,000 and over" class was $20,000 and that the class average for the "$25,000 and over" class was $44,000. (The former applies to 1950 and 1960 sectors, and the latter applies to 1960 All.) These figures were based on application of the Pareto distribution. Cf., Herman P. Miller, Trends in the Income of Families and Persons in the United States: 1947 to 1960, U. S. Bureau of the Census Technical Paper No. 8, 1963, p. 16.

2/ U. S. Department of Commerce, Office of Business Economics, Survey of Current Business, for the years 1950, 1951, and 1953 through 1959.

and by substitution

$$\frac{Y_{mi}}{Y_{ai}} = C_o + C_1 (S_{mi})$$

where

Y_{mi} = the first moment (or mean income estimate) computed under the midpoint assumption

Y_{ai} = the first moment computed using each interval's average income

S_{ai} = the third deviation moment computed using each interval's average income

and

S_{mi} = the third deviation moment computed under the midpoint assumption.

For purposes of estimation, both sides of the equation are multiplied by Y_{ai}; and a disturbance term, U_i, is introduced to yield:

$$Y_{mi} = C_o Y_{ai} + C_1 (S_{mi}) (Y_{ai}) + U_i.$$

Results obtained for the three cases (farm operator family, nonfarm family, and all family) are listed in Appendix Table A-2.

To measure the "goodness of fit," the coefficient of multiple determination was computed for each case. For all three cases, this coefficient was at least .99. The standard error of the dependent variable was small in each of the three cases. For the farm operator family, nonfarm family, and all family cases, the standard errors were \$6.62, \$3.52, and \$3.75, respectively.

Given the estimates of C_o and C_1 from the fitted equation, averages for particular income distributions were obtained using:

APPENDIX TABLE A-2

Estimated Coefficients for Average Family Income as a Function of Midpoint Estimate and Skewness

Family sectors	Estimated coefficients	
	Midpoint estimate $\hat{C}_o$ a/	Midpoint estimate times skewness measure $\hat{C}_1$ a/
Farm operator b/	0.9927	$1.0435\ 10^{-14}$
Nonfarm	1.0016	$0.9140\ 10^{-15}$
All	1.0043	$7.9020\ 10^{-15}$

a/ Estimates are indicated by the hat (^).

b/ The results for farm operator families were assumed to be applicable to the rural farm family units in this study.

Sources:

U. S. Department of Commerce, Office of Business Economics, Survey of Current Business, Vol. 43, No. 7 (July, 1963), p. 22.

Idem, U. S. Income and Output: A Supplement to the Survey of Current Business, 1958, p. 161.

Idem, Income Distribution in the United States: A Supplement to the Survey of Current Business, 1953, p. 85.

$$Y_{ai} = \frac{Y_{mi}}{(\hat{C}_o + \hat{C}_1 S_{mi})}$$

and then inserting values of Y_{mi} and S_{mi} obtained for the particular distribution.

For the state as a whole, midpoint estimates were fairly close to the adjusted averages estimated here. Thus, for the family and unrelated unit in 1950, the ratio of the adjusted average to the midpoint average was .986 for the nonfarm sector and 1.065 for the farm sector. On a county level, however, more pronounced differences occurred.

Underreporting of Income and Imputed Income

Census data differs from OBE data on income; the latter include the familiar personal income and national income series issued by the federal government.

Appendix Table A-3 compares aggregate income estimates for the state obtained from the 1950 and 1960 Census data to the OBE data for the years 1949 and 1959, respectively.[1/]

It turns out that the ratio of this study's estimated aggregate income to that of the OBE is .80 in 1950 and .90 in 1960. Three factors account for this difference: (1) imputed income for owned housing and homegrown food does not appear in the Census figures but does appear in the OBE figures; (2) the value of services of banks and the income of persons who died or emigrated appears in the OBE but not in the Census figures; and (3) some income is underreported in the Census data, particularly dividend and interest income. Estimates for 1950 of the impact of each of these factors in terms of understatement of OBE figures consist of the following: .949 for omission of imputed income, .956 for exclusion of bank services and income of persons who died or emigrated,

1/ *Ibid*., Vol. 45, No. 7 (July, 1965), p. 10.

APPENDIX TABLE A-3

Comparison of Aggregate Family Income Estimates, 1950 and 1960

	Aggregate family income estimates	
	1950[a]	1960[a]
	million dollars	
Census		
Midpoint estimates	14.285[b]	36.888[b]
Present study averages	14.235[b]	36.814[c]
Office of Business Economics	17.835	40.960
	Ratio	
Present study averages relative to Office of Business Economics	.798	.899

a/ Year refers to date of Census. Dates of income received are 1949 and 1959, respectively.

b/ Aggregate figures were obtained by multiplying average estimates for All units by total number of income units as listed in the Census of Population.

c/ Estimated by multiplying midpoint estimate times .998, the ratio of this study estimates to midpoint estimates for 1960 All family income.

Sources:

U. S. Bureau of the Census, U. S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 5, California, 1952, p. 5-71.

Idem, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963, p. 6-250.

U. S. Office of Business Economics, Survey of Current Business, Vol. 45, No. 1 (July, 1965), p. 10.

and .885 reflecting underreporting of income.[1] The product of these fractions is .803, approximately that obtained as the overall ratio here. The increase of the ratio to .90 in 1959 probably reflects less imputed income from home-grown food and increased accuracy of measurement of income from dividends and interest.

If income figures corresponding to OBE measures are desired, the ratios presented in Appendix Table A-3 can be used as adjustment factors.

Family Income Versus Family and Unrelated Income

Comparison of farm and nonfarm sectors using family income does not necessarily correspond to a comparison of sectors using family and unrelated income. This is because the relative proportions of families and unrelated individuals differ between the rural farm and the nonfarm sectors. The average income of unrelated individuals is well below that of families so that comparisons of farm and nonfarm income are affected by choice of income unit. Appendix Table A-4 exhibits the fraction of families and of unrelated individuals in the family and unrelated category by sector for the 1950 and 1960 Census.

1/ Hoch, op. cit. pp. 7, 30, and 31.

APPENDIX TABLE A-4

Fraction of Families and Fraction of Unrelated Individuals in the Family and Unrelated Unit by Sectors, 1950 and 1960

Year and sectors	Family and unrelated	
	Number of families as fraction of families and unrelated individuals total	Number of unrelated individuals as fraction of families and unrelated individuals total
1950		
All	.720	.280
Urban	.717	.283
Rural nonfarm	.717	.283
Rural farm	.787	.213
1960		
All	.718	.282
Urban	.719	.281
Rural nonfarm	.688	.312
Rural farm	.851	.149

Sources:

U. S. Bureau of the Census, U. S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 5, California, 1952, p. 5-71.

Idem, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963, p. 6-250.

APPENDIX B

County Ordering and County Income Rank

County Ordering

Counties were ordered on the basis of location running from north to south and are listed in the text tables on the basis of that order. Figure B-1 exhibits a map of California counties coded in the location order employed here and presents an alphabetical list of counties with corresponding location code number.

County Rank by Average Income

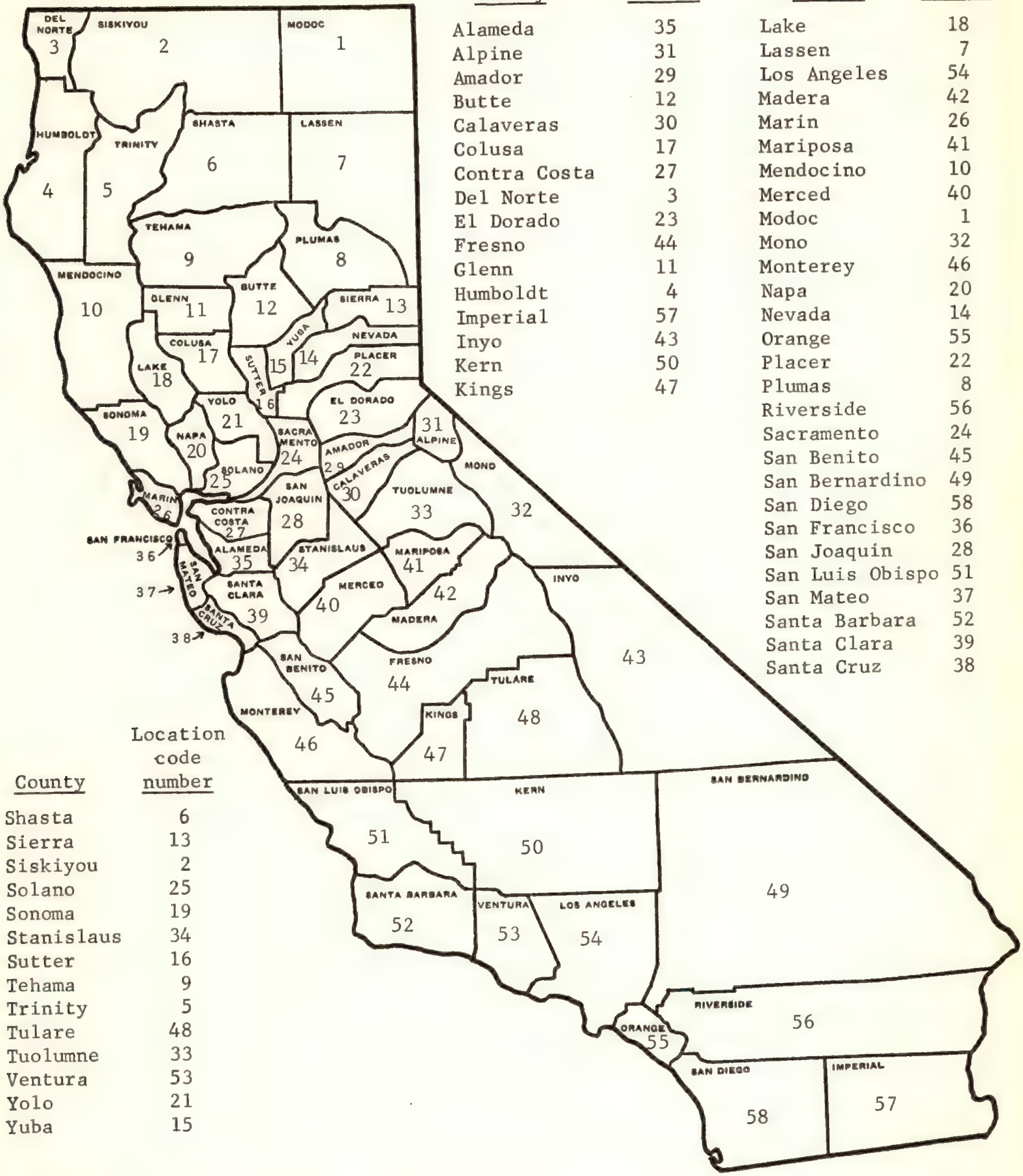
The average income figures of Tables 2 and 3 were used to develop county rankings. Appendix Table B-1 presents nonfarm and rural farm rankings for 1950 and 1960 and then exhibits the change in rankings defined as the 1950 rank minus the 1960 rank. Data limitations pose something of a problem. Thus, the 1950 data is for family and unrelated income while the 1960 data is for family income. Again, some counties appearing in 1950 are omitted in 1960. For the 1950 rural farm sector (using family and unrelated data), the first five counties in rank order were Alpine, Ventura, Colusa, Orange, and Glenn. For 1960 (using family data), the first five in rank order were Orange, San Mateo, Ventura, Del Norte, and Los Angeles. The lowest five counties for 1950 were Placer, Lake, Del Norte, Imperial, and Trinity, with Trinity the lowest. In 1960 the lowest five counties (of those reported) were Stanislaus, Tuolumne, Mariposa, Calaveras, and Mono.

Rural farm income tends to be more variable than nonfarm income so that greater change in rankings can be expected and in fact occurs. Such change is particularly pronounced for counties with small populations, for example, Del Norte and Alpine counties.

The information in Appendix Table B-1 is the source for Figures 4 and 5 presented in the text.

County	Location code number
Alameda	35
Alpine	31
Amador	29
Butte	12
Calaveras	30
Colusa	17
Contra Costa	27
Del Norte	3
El Dorado	23
Fresno	44
Glenn	11
Humboldt	4
Imperial	57
Inyo	43
Kern	50
Kings	47

County	Location code number
Lake	18
Lassen	7
Los Angeles	54
Madera	42
Marin	26
Mariposa	41
Mendocino	10
Merced	40
Modoc	1
Mono	32
Monterey	46
Napa	20
Nevada	14
Orange	55
Placer	22
Plumas	8
Riverside	56
Sacramento	24
San Benito	45
San Bernardino	49
San Diego	58
San Francisco	36
San Joaquin	28
San Luis Obispo	51
San Mateo	37
Santa Barbara	52
Santa Clara	39
Santa Cruz	38



County	Location code number
Shasta	6
Sierra	13
Siskiyou	2
Solano	25
Sonoma	19
Stanislaus	34
Sutter	16
Tehama	9
Trinity	5
Tulare	48
Tuolumne	33
Ventura	53
Yolo	21
Yuba	15

FIGURE B-1. Alphabetical List of Counties with Corresponding Location Code Numbers

APPENDIX TABLE B-1

County Rankings of Average Income and Changes in Ranks Over Time
1950 and 1960

County	Rankings				Change in rankings	
	1950 family and unrelated income		1960 family income		1950 rank minus 1960 rank	
	Nonfarm	Rural farm	Nonfarm	Rural farm	Nonfarm	Rural farm
Modoc	21	31	36	14	-15	17
Siskiyou	15	7	43	38	-28	-31
Del Norte	50	55	19	4	31	51
Humboldt	6	13	16	11	-10	2
Trinity	57	57	18	a/	39	b/
Shasta	18	37	25	35	- 7	2
Lassen	7	15	31	37	-24	-22
Plumas	11	9	28	41	-17	-32
Tehama	31	19	46	47	-15	-28
Mendocino	32	48	29	46	3	2
Glenn	28	5	50	45	-22	-40
Butte	40	28	48	30	- 8	- 2
Sierra	41	49	40	a/	1	b/
Nevada	54	52	44	29	10	23
Yuba	38	45	51	40	-13	5
Sutter	23	27	30	23	- 7	4
Colusa	17	3	38	10	-21	- 7
Lake	56	54	56	33	0	21
Sonoma	39	38	33	39	6	- 1
Napa	27	35	15	9	12	26
Yolo	33	30	17	13	16	17
Placer	26	53	21	44	5	9
El Dorado	49	39	12	26	37	13
Sacramento	3	41	6	12	- 3	29
Solano	22	50	20	31	2	19
Marin	2	20	1	16	1	4
Contra Costa	5	8	4	8	1	0
San Joaquin	24	46	26	24	- 2	22

(Continued on next page.)

APPENDIX TABLE B-1--continued.

County	Rankings				Change in rankings	
	1950 family and unrelated income		1960 family income		1950 rank minus 1960 rank	
	Nonfarm	Rural farm	Nonfarm	Rural farm	Nonfarm	Rural farm
Amador	42	40	39	15	3	25
Calaveras	48	47	35	52	13	- 5
Alpine	58	1	58	a/	0	b/
Mono	43	6	11	a/	32	b/
Tuolumne	36	26	41	50	- 5	-24
Stanislaus	35	25	49	49	-14	-24
Alameda	4	16	10	21	- 6	- 5
San Francisco	10	c/	9	c/	1	c/
San Mateo	1	29	2	2	- 1	27
Santa Cruz	44	42	42	36	2	6
Santa Clara	12	23	3	17	9	6
Merced	30	24	53	48	-23	-24
Mariposa	55	51	57	51	- 2	0
Madera	52	36	55	42	- 3	- 6
Inyo	16	14	45	a/	-29	b/
Fresno	19	34	32	34	-13	0
San Benito	45	32	47	18	- 2	14
Monterey	29	33	24	27	5	6
Kings	14	17	52	43	-38	-26
Tulare	47	22	54	20	- 7	2
San Bernardino	46	44	23	28	23	16
Kern	8	18	22	25	-14	- 7
San Luis Obispo	51	12	37	22	14	-10
Santa Barbara	20	11	8	6	12	5
Ventura	13	2	14	3	- 1	- 1
Los Angeles	9	10	7	5	2	5
Orange	2	4	5	1	20	3
Riverside	58	43	27	32	26	11
Imperial	37	56	34	19	3	37
San Diego	34	21	13	7	21	14

(Continued on next page.)

APPENDIX TABLE B-1--continued.

a/ No data available.

b/ At least one year not available. This was more pronounced for 1960, so there is some tendency for an upward shift in rankings in 1960. Ranks appearing in 1950 but not in 1960 are: 53, 54, 55, 56, and 57. Some counties do not appear in 1960. Their ranks in 1950 were 1, 14, 49, 55, and 57.

c/ Not applicable.

Sources: Developed from data in U. S. Bureau of the Census. U. S. Census of Population: 1950. Vol. II, Characteristics of the Population, Part 5, California, 1952, p. 5-71; and idem, U. S. Census of Population: 1960. Vol. I, Characteristics of the Population, Part 6, California, 1963, p. 6-250.

APPENDIX C

Some Univariate Results for the Income Determination Model

In the discussion of the multivariate model of factors determining the income distribution, there was a focus on the set of equations involved, viewed as a system. For example, there were statistical tests of hypothesized zero coefficients for a given independent variable in all equations. Some univariate results are presented here, both for general interest and for possible contrasts and comparisons.

Calculated t Statistics

Appendix Tables C-1 and C-2 present calculated t statistics for the 1950 and 1960 equations, respectively. (Though these test statistics are applicable in multivariate models, they involve the consideration of one variable at a time in a particular equation. Because this differs from the focus adopted in the text, it seemed proper to catalogue these statistics as univariate results.) The number of t values significant at the 5 percent level ranges from a low of two for irrigation and for tenure and to a high of nine for education, viewing the 1950 and 1960 results in combination. All variables except irrigation and tenure have at least five significant coefficients. Given a true coefficient value of zero, the 25 cases examined could be expected to yield only one such result per variable.

Regressions Using Average Income

Appendix Table C-3 exhibits regression results obtained by regressing average income on the independent variables employed here. In this case, of course, there is only one equation per year. The R^2 values here are .46 and .64 for 1950 and 1960, respectively. This contrasts with an explained variance of around .99

APPENDIX TABLE C-1

Calculated t-Ratios, 1950

Constant term and independent variables	Income class interval[a]						
	1	2	3	4	5	6	7
	coefficient estimates						
Constant term	1.780	0.881	0.713	1.672	2.294	0.931	-0.372
Education	-2.558	-4.320	-2.328	-1.528	0.939	0.850	0.153
Road Class 1	-2.511	-0.725	-0.884	1.301	1.269	2.030	1.124
Tenure (managers)	0.714	0.071	0.337	1.330	2.573	-0.616	1.301
Age level	0.244	1.222	0.698	-0.436	-1.563	-0.688	0.560
Trading center	-1.828	0.835	2.976	2.832	1.585	0.432	-0.994
Irrigated farms	-1.140	-0.119	0.887	-0.514	-1.535	0.315	0.791
Farm size	-2.013	-1.075	-2.650	-0.341	-0.617	0.568	2.281
Real estate	-1.302	-0.865	-0.769	-2.400	-1.206	1.397	-1.493
Urban index	-1.048	-0.594	-1.899	-2.183	-2.075	-2.092	1.073
Isard index	0.565	-0.356	1.370	1.797	0.339	2.155	-1.216
	8	9	10	11	12	13	14
Constant term	-1.259	1.523	-1.389	0.706	-1.640	1.293	-0.929
Education	1.897	2.855	-0.910	1.288	0.484	1.356	1.735
Road Class 1	-1.397	0.364	1.344	0.944	-2.620	-1.819	0.405
Tenure (managers)	0.296	-1.957	0.206	-2.497	-1.352	-0.864	-0.480
Age level	1.338	-2.568	2.071	-1.522	1.964	-1.031	0.675
Trading center	-1.776	0.813	-1.226	0.488	-1.311	-3.109	-0.031
Irrigated farms	1.362	-1.115	0.079	-0.578	2.488	0.926	-0.373
Farm size	1.134	-0.091	1.386	-0.786	0.654	1.754	0.615
Real estate	-0.732	2.318	-0.742	2.210	1.270	0.464	2.439
Urban index	2.263	1.752	-0.636	1.661	2.238	1.380	0.808
Isard index	-3.210	1.472	-0.202	2.338	-2.148	-1.382	-2.663

a/ See supra, Table 1, page 6, for income values included in each class interval.

APPENDIX TABLE C-2

Calculated t-Ratios, 1960

Constant term and independent variables	Income class interval[a/]					
	1	2	3	4	5	6
	coefficient estimates					
Constant term	1.770	0.696	0.789	1.904	-0.074	-1.886
Education	-0.206	-3.035	-3.118	-1.514	-2.610	-0.246
Road Class 1	-0.973	-1.822	0.504	1.326	2.604	1.541
Tenure (managers)	-1.558	0.278	0.936	-1.832	0.442	0.641
Age level	-0.642	1.438	0.586	-1.202	1.068	1.990
Trading center	-0.814	0.569	1.352	1.286	0.312	1.041
Irrigated farms	-1.298	0.351	0.904	0.040	-0.416	2.144
Farm size	-0.523	1.662	0.174	-2.280	3.127	2.786
Real estate	-0.726	-2.724	-2.539	1.966	-1.786	-0.723
Urban index	-0.359	-2.303	-0.163	-0.332	0.058	0.480
Isard index	-1.393	1.064	-0.316	1.705	0.221	-1.466

(Continued on next page.)

APPENDIX TABLE C-2--continued.

Constant term and independent variables	Income class interval[a]				
	7	8	9	10	11
	coefficient estimates				
Constant term	-0.917	-0.120	0.199	0.670	1.315
Education	-0.977	-0.359	0.984	3.030	4.962
Road Class 1	-0.258	1.418	-0.337	-2.523	-1.426
Tenure (managers)	0.461	1.580	-0.764	0.571	-0.199
Age level	1.313	0.648	0.138	-0.822	-2.326
Trading center	1.748	1.352	0.524	-1.275	-2.135
Irrigated farms	0.436	0.175	-1.072	0.068	-0.815
Farm size	-0.074	2.716	-1.031	-0.243	-3.101
Real estate	-0.925	-2.540	0.930	-0.021	4.523
Urban index	1.201	-0.728	-0.199	0.824	0.899
Isard index	-0.948	0.422	-1.277	-2.140	1.439

a/ See supra, Table 1, page 6, for income values included in each class interval.

APPENDIX TABLE C-3

Results Obtained in the Regression of Average Income
1950 and 1960

Constant term and independent variables	Coefficients		t-ratios	
	1950[a]	1960[b]	1950	1960
Constant term	-0.722	9.018	-0.273	1.516
Education	0.293	0.956	3.471	5.265
Road Class 1	0.547	-1.339	1.242	-1.223
Tenure (managers)	-5.834	1.130	-0.760	0.156
Age level	0.092	-2.353	0.178	-1.990
Trading center	-0.039	-0.150	-1.154	-2.049
Irrigated farms	0.369	-0.432	1.403	-0.658
Farm size	0.509	-1.002	2.048	-2.591
Real estate	0.010	0.012	2.184	4.222
Urban index	0.533	2.149	0.737	1.152
Isard index	-0.577	0.565	-1.563	0.898

a/ R^2_{1950} = .4628.

b/ R^2_{1960} = .6391.

of total variance in the multivariate case as shown in Table 15. In the univariate case, t values significant at the 5 percent level are obtained for education, farm size, and real estate in both 1950 and 1960 and for trading center in 1960. However, the farm size coefficient changes sign between those years. The urban index, which was an important explanatory variable in the 1950 multivariate model (based on Tables 15 and 16) does not have a significant t value in either 1950 or 1960 for this univariate case.

In general, of course, the univariate case yields much less information than does the multivariate case.